

(TM)

\*\*\*\*\*

Distribution rights by Oxford Molecular Ltd

09/07/2018

aaaa 865

tttt  
aaaa 000

777

13

5

9759

525

11

77e-10

9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

# FEATURES

Location/Qualifiers  
1..496

/organism="Homo sapiens"

/notes="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

/db\_xref="ATCC (inhost):157086"

/db\_xref="taxon:9606"

/clone\_lib="Jurkat T-cells v"

/cell\_type="T-lymphocyte"

<1..5496

98 a 143 c 151 g 103 t 1 others

Query Match 53.5%; Score 463; DB 24; Length 496;

Best Local Similarity 99.2%; Pred. No. 2.06e-287;

Matches 482; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Db 5 TTTGTGGAGGCGCAGAGAGTACCAGCTGGACATCTTCTGCTGAGAGCCCCAGG 64

QY 1 ttgtgagggcagcagagagatccacagctggacatcttctctgtatgagcccccag 60

Db 65 CTGAGGTGCCCTGCTCATATGCTCTCCCGAGGCTCTCGCCCTGGCGAGGTCATCAT 124

QY 61 ctggaggtgcccgtctcacatgctcttccccaggctctgcctggcaggtcatcata 120

Db 125 GTACGGGAGCTGGTCTTCACAGAGCCGGAAGCATTTTACTGTGAGCCTGAGGACCAAGCT 184

QY 121 gtacggggagctggtcttgcaagagccgaagcatttactgtgagcctgagggaccaggt 180

Db 185 GCCCATGCTCTGTGACACTCAGGSCCTCTTCGACAGACAGACTCTGCCCTGTCTCC 244

QY 181 gccatgctctctgtgacactcagggcctctcttcgacagacaaactctggctggatctcc 240

Db 245 CGCTGGGGGCGAGAGAACTGATCTACGCCCTCTTCCTCTTTTACCCACAGAGATCTTT 304

QY 241 cgctgggggagagaaactgatctcagccctctctcttttaccctcagagattcttt 300

Db 305 GAGGTGCTGCTCTTCCAGAGAGGAGGTGAAGCTGGCGCTCAATGGCGAGGGGCTG 364

QY 301 gagggtgctcctgttccagagggagggctgaagctggcgctcaatggcgagggctg 360

Db 365 GGGCCACACGATGAACACGAGCCCTGACAGAGCTGCGGAGCTCCGGATCAGTGA 424

QY 361 ggggcccacagcagatgaacacagcagccctggagcagctgcggagctcgcgagtcagtgga 420

Db 425 AGTGTCAGCTCTACTGTGTCCACTCTCTGA-GGATGGTTCCAGGGAATACCGNAGAA 483

QY 421 agtgccagctcactgtgtccactctggaaggatggtttccagg-aaataccg-cagaaa 478

Db 484 ACAGA 489

QY 479 acaaga 484

# RESULT

LOCUS H30148 411 bp mRNA EST

DEFINITION y059h05.r1 Homo sapiens cDNA clone 182265 5'

ACCESSION H30148

NID 9901058

KEYWORDS EST.

SOURCE human clone-182265 library=Soares breast 3NHBst vector=PT7T3D

(Pharmacia) with a modified polylinker host=DH10B (ampicillin

resistant) primer=M13Rpl Rsitel=Not I Rsitel=Not I Eco RI Advt human.

1st strand cDNA was primed with a Not I - oligo(dT) pr mer [5'

TGTTACCAATCTGAAGTGGAGCGGCGCCCTTTTCTTTTTTTTTTTT 3']

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo.

# ORGANISM

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;

Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 411)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 264

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Location/Qualifiers

1..411

/organism="Homo sapiens"

/clone="182265"

90 a 113 c 121 g 79 t 8 others

BASE COUNT

ORIGIN

Query Match 38.8%; Score 336; DB 16; Length 411;

Best Local Similarity 95.7%; Pred. No. 2.64e-199;

Matches 360; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

Db 1 CAGGCTCTCTNAACAGGCGCAGGTGATCATAGTACGGGAGTGTCTTNCAGAGCCGAG 60

QY 91 cagggtctctcgtcgtggcaggtcatcatgtagcagggaggtggtcttgcagagccgag 150

Db 61 CATTTTACTGTGAGCCTGAGGACGACGAGCTNCCATGTCTCTGTGACACTCAGGCGCTCC 120

QY 151 cattttactgtgagcctgagggaccaggtgcccagtcctctgtgacactcagggcctcc 210

Db 121 TTCGACAGACAACTCTGGCCTGGATCTCCCGCTGGGGCAGAGAAACTGATCTCAGCC 180

QY 211 ttgcgacagaaactctggcctggatctcccgctggggcagagaaactgctcagcc 270

Db 181 CCCTTCCTCTTTTACCNCAGAGATCTTTGAGGTGCTGCTCCTCTTCCAGAGGGGG 240

QY 271 cctctcctcttttaccncagagattctttgaggtgctgctcctgttccagagggaggg 330

Db 241 CTGAAGCTGGCGCTCAATGGGCGAGGCTGGGGGCCACCAAGCAATGAACACGAGGCC 300

QY 331 ctgaagctggcgctcaatggggcaggggctggggggccacca-gca-tgaaccgagggccc 388

Db 301 TGGAGCAGCTGCGNGACTCCGGATCAGTGAAGTATCCAGCTCTACTGTGTCCACTCTCT 360

QY 389 tggagcagctgaggagctccgagcagtggaagtgctcagctctactgtgtccactctct 448

Db 361 NA-GGATGTTCCAGG 375

QY 449 gaaggtggttccagg 464

RESULT 3 H27468 422 bp mRNA EST

LOCUS y059h05.s1 Homo sapiens cDNA clone 182265 3'

DEFINITION

\*\*\*\*\*

W P E R L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Dec 15 15:55:11 1998; MasPar time 5.67 Seconds  
629.394 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-109-864-2  
Description: (1-133) from US09109864.pep  
Perfect Score: 980  
Sequence: 1 MSPRLVPCSHALPQGLSPG.....LEQLRELISGSVOLYCVHS 133  
Scoring table: PAM 150  
Gap 11  
Searched: 74019 seqs, 26840295 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot35  
1:swissprot  
Statistics: Mean 44.065; Variance 78.778; scale 0.559

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	174	17.8	144	1	LEG5_RAT	2.71e-14
2	146	14.9	354	1	LEG9_RAT	2.89e-09
3	140	14.3	316	1	LEG8_HUMAN	3.11e-08
4	140	14.3	353	1	LEG9_MOUSE	3.11e-08
5	136	13.9	135	1	LEG7_HUMAN	1.49e-07
6	125	12.8	355	1	GALECTIN-9 (HOM-HD-21)	9.88e-06
7	112	11.4	279	1	LE32_CAEEL	1.12e-03
8	112	11.4	289	1	LE33_CAEEL	1.12e-03
9	109	11.1	316	1	LEG8_RAT	3.21e-03
10	109	11.1	324	1	LEG4_RAT	3.21e-03
11	106	10.8	448	1	GPT_YEAST	9.03e-03
12	99	10.1	134	1	LEG4_CHICK	9.41e-02
13	98	10.0	308	1	NOSF_PSEST	1.30e-01
14	97	9.9	184	1	YRNA_CAEEL	1.80e-01
15	95	9.7	746	1	PFPA_PSEAE	3.42e-01
16	94	9.6	135	1	LEG7_RAT	4.70e-01
17	93	9.5	493	1	GSPE_ECOLI	6.43e-01
18	93	9.5	1522	1	CINL_LOLBI	8.79e-01
19	92	9.4	437	1	SECY_STRGR	8.79e-01
20	92	9.4	583	1	CFAL_HUMAN	8.79e-01
21	92	9.4	620	1	KEFC_ECOLI	8.79e-01
22	90	9.2	510	1	EMRB_HAEIN	1.63e+00
23	90	9.2	761	1	POOF_KLEPN	1.63e+00

24	89	9.1	225	1	NTCA_SYNY3	GLOBAL NITROGEN REGULA	2.21e+00
25	89	9.1	323	1	LEG4_HUMAN	GALECTIN-4 (LACTOSE-BI	2.21e+00
26	89	9.1	563	1	YEOK_SCHPO	HYPOTHETICAL PROTEIN C	2.21e+00
27	88	9.1	976	1	FIBP_ADEB3	FIBER PROTEIN.	2.21e+00
28	88	9.0	137	1	YSCB_YEREN	PUTATIVE YOP PROTEINS	2.98e+00
29	88	9.0	218	1	SIR_DESVH	SULFITE REDUCTASE, ASS	2.98e+00
30	87	8.9	198	1	COAT_BSVV	COAT PROTEIN (CAPSID P	4.02e+00
31	87	8.9	262	1	SPRE_RAT	SEPIAPTERIN REDUCTASE	4.02e+00
32	87	8.9	343	1	HOLA_ECOLI	DNA POLYMERASE III, DE	4.02e+00
33	87	8.9	786	1	YBIO_ECOLI	HYPOTHETICAL 86.8 KD P	4.02e+00
34	87	8.9	1513	1	RPOD_ORYSA	DNA-DIRECTED RNA POLYM	4.02e+00
35	87	8.9	1527	1	RPOD_MATZE	DNA-DIRECTED RNA POLYM	4.02e+00
36	87	8.9	1742	1	GUNA_CALSA	ENDOGLUCANASE A PRECUR	4.02e+00
37	86	8.8	481	1	ATZB_PSED	HYDROXYLASE HYDROL	5.40e+00
38	85	8.7	223	1	NTCA_ANASP	GLOBAL NITROGEN REGULA	7.24e+00
39	85	8.7	244	1	LEG3_CRILLO	GALECTIN-3 (GALACTOSE-	7.24e+00
40	85	8.7	424	1	MTBL_BACSH	MODIFICATION METHYLASE	7.24e+00
41	85	8.7	495	1	ACHP_RAT	NEURONAL ACETYLCHOLINE	7.24e+00
42	85	8.7	725	1	GUNG_CLOCE	ENDOGLUCANASE G PRECUR	7.24e+00
43	85	8.7	868	1	MCE_ASF7	MRNA CAPPING ENZYME (M	7.24e+00
44	85	8.7	879	1	GUNI_CLOTM	ENDOGLUCANASE I PRECUR	7.24e+00
45	85	8.7	1108	1	MAN2_DROME	ALPHA-MANNOSIDASE II (	7.24e+00

ALIGNMENTS

RESULT 1  
ID LEG5\_RAT STANDARD; PRT; 144 AA.  
AC P47967;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE GALECTIN-5 (RL-18).  
GN LGALS5.  
OS RATIUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-RETICULOCYTES;  
RX MEDLINE; 95197487.  
RA GITT M.A., WISER M.F., LEFFLER H., HERRMANN J., XIA Y.-R.,  
RA MASSA S.-M., COOPER D.N.W., JUSIS A.-J., BARONDES S.H.,  
RA J. BIOL. CHEM. 270:5032-5038(1995).  
CC -!- FUNCTION: MAY FUNCTION IN ERYTHROCYTE DIFFERENTIATION.  
CC -!- SUBUNIT: MONOMER.  
CC -!- TISSUE SPECIFICITY: ERYTHROCYTES.  
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
DR EMBL; L21711; G785053;  
DR PROSITE; PS00309; LECTIN GALACTOSIDE; 1. QP801.57  
KW GALAPTIN; LECTIN; ACETYLATION.  
FT INIT\_MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT BINDING 76 82 BETA-GALACTOSIDE (POTENTIAL).  
FT CONFLICT 127 135 MISSING (IN G785053).  
SQ SEQUENCE 144 AA; 16065 MW; 5C7CC533 CRC32;

Query Match 17.8%; Score 174; DB 1; Length 144;  
Best Local Similarity 28.0%; Pred. No. 2.71e-14;  
Matches 37; Conservative 38; Mismatches 49; Indels 8; Gaps 6;  
Db 10 PNLAVFFTSIPNGLVPSKIVISGVVLSDAKRFQNLRC-GGDIAPHLNPRDENAVR 68  
QY 3 PRLEVPCHALPQGLSPGVIIIVRLVQPKFTVSLRQAAHAPVTLRASFADRTLAW 62  
Db 69 NTOINNSWGPERSLSPMPFSRGORF-SWVILCEGHCFFKAVDVGQHICYSRLNLPD 127  
QY 63 ---ISR-WG-QKKLISAPFLFP-QRFEEVLLLFEGGKLALNGOGLGATSMNQALEQ 116  
Db 128 INFLEVAGDIQL 139  
QY 117 LRELISGSVOL 128

```

RESULT      3
ID ID LG8HUMAN STANDARD; PRT; 316 AA.
AC O00214;
DT 01-NOV-1997 (REL. 35, CREATED)
DT DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DT DE GALECTIN-8 (PROSTATE CARCINOMA TUMOR ANTIGEN) (PCTA-1).
GN LGAIS9.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [ ]
RP SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RM MEDLINE; 96293510.
LA SU Z.-Z., LIN J., SHEN R., FISHER P.E., GOLDSTEIN N.I., FISHER P.B.;
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLY);
CC -! TISSUE SPECIFICITY: SELECTIVE EXPRESSION BY PROSTATE CARCINOMAS
CC VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC HYPERTROPY.
CC -! DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -! SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; L78132; GI932712; -.
DR DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW KW GALAPTIN; LECTIN; REPEAT; ANTIGEN.
FT FT DOMAIN 1 153 GALAPTIN 1.
   FT DOMAIN 154 184 LINKER.
FT FT DOMAIN 185 315 GALAPTIN 2.
FT BT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 35588 MW; EB4D9A3D CRC32;

Query Match          14.3%; Score 140; DB 1; Length 316;
Best Local Similarity 24.6%; Pred.No. 3.1le-08;
Matches    34; Conservative 41; Mismatches 56; Indels       7; Gaps

Db D179 TPQSLPSFAARLNTPMGPRGTVVVGVEVNANAKSFNVDLGAGSKDIALLHNRLNIKAF 238
   QY :|::||:||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
     2 SPRLEVPCHALPGQLSPGVGIIVGLVLQEPKHVTLSL-RDQAHAHPVTLRASFADRTL 60

Db D239 VRNSELQESGEERNITSPFPSPGWTFEMIIICYDVREFKVAVNGVHSLEYKKRFKE-L5 297
   QY :|::||:||::||::||::||::||::||::||::||::||::||::||::||::||::||:
     61 AWIS--R-WGQKLKSAPFLFYQPREFVFLLLFQEGLXIALNG-QGLGATSMNQOALE 115

Db D298 SIDTLEIWDGHILEVERS 315
   QY |::||::||::||::||::||::||:
     116 QLRELRISGSVOLYCVDHS 133


RESULT      4
ID ID LEG9.MOUSE STANDARD; PRT; 353 AA.
AC O08573; O08572;
DT 01-NOV-1997 (REL. 35, CREATED)
DT DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT DE GALECTIN-9.
GN LGAIS9.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [ ]
RP SEQUENCE FROM N.A.
RX SPRAIN-CD-1; TISSUE=SMALL INTESTINE, AND KIDNEY;
RM MEDLINE; 97190351.
LA WADA J., KANWAR Y.S.;
JL J. BIOL. CHEM. 272:6086(1997).
CN [2]
RR CHARACTERIZATION.
RX MEDLINE; 97298141.
RL WA DA J., OTA K., KU MAR A., WALLNER E.I., KANWAR Y.S.:
CC -! FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
    EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
```

```
CC -!- TISSUE SPECIFICITY: MAINLY IN STRATIFIED SQUAMOUS EPITHELIUM.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; L07769; G182132; -.
DR EMBL; U06643; G458703; -.
DR AARHUS/GHEHT-2DPAGE; 17; IEF.
DR MTM; 600615; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW GALAPTIN; LECTIN.
FT INIT MET      0
FT BINDING       69   75   BETA-GALACTOSIDE (POTENTIAL).
SQ SEQUENCE     135 AA; 14944 MW; 33656FP3 CRC32;

Query Match      13.9%; Score 136; DB 1; Length 135;
Best Local Similarity 28.1%; Pred. No. 1.49e-07;
Matches          38; Conservative 39; Mismatches 47; Indels 11; Gaps

Db      2 NVPHKSLPEGIRPGVIVLRIGLVPPNASEFHVNLGCEQSDAALHNPRLDISEVF 61
        :|||:::||:|||||:::|||||:::|||||:::|||||:::|||||:::
QY      6 EVPCSHALPGLSGGVIVRGLVPEKHFTVSL---RDQAHAVPVTLRASFDRTLAW 62
        :|||:::||:|||||:::|||||:::|||||:::|||||:::|||||:::
Db      62 NSREQSGWSEERGPGVPFOR-QGPT-EVLIIASDDGFK-AVVGDQAYHHFRRLPLARV 118
        |:|::||:|||||:::|||||:::|||||:::|||||:::|||||:::
QY      63 ISR----WG-QKKLIISAPFLFYQRFEFVILLFQEGLKLALNGOGLGATSNMQALEOL 117
        |:|::||:|||||:::|||||:::|||||:::|||||:::|||||:::
Db      119 RUVVEGVDVOLDNR 133
        |:|::||:|||||:::|||||:::|||||:::|||||:::|||||:::
QY      118 RELRISSGVOLYCVH 132
```

RESULT	6		
ID	LEG9_HUMAN	STANDARD;	PRT; 355 AA.
AC	000182;	014532;	
DT	01-NOV-1997	(REL. 35, CREATED)	
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)	
DE	GALECTIN-9 (HOM-HD-21).		
GN	LGALS9.		
OS	OS HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; PRIMATES.		
	[1]		
RP	SEQUENCE FROM N.A. (SHORT FORM).		
RC	TISSUE=SPLEN.		

RA	TUERICI O., SCHMITT H., FADLE N., PFREUNDSCUH M., SAHIN U.;
RJ	J. BIOL. CHEM. 272:6416-6422(1997).
[2]	
RN	SEQUENCE FROM N.A. (LONG FORM).
RP	TISSUE-GASTRIC CARCINOMA;
RC	KATO S.;
RA	SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC	-!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
CC	-!- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC
CC	TISSUES; OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.
CC	-!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC	DOMAINS.
CC	-!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
CC	-!- ALTERNATIVE SPLICING.
CC	-!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR	EMBL; Z49107; E310691; -
DR	EMBL; AB006782; D1023026; --
DR	MIM: 601879; -
DR	PROSITE; PS00309; LECTIN_GALACTOSIDE; 2.
KW	GALAPTIN; LECTIN; REPEAT; ALTERNATIVE SPLICING.
FT	DOMAIN 1 148
FT	DOMAIN 149 206
FT	DOMAIN 207 355
FT	BINDING 82 88
FT	BINDING 287 293
FT	VARSPLIC 149 180
FT	CONFLICT 88 88
FT	CONFLICT 135 135
FT	CONFLICT K -> R (IN REF. 1).
FT	CONFLICT S -> P (IN REF. 1).

```

01-OCT-1996 (REL. 34, CREATED)
01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
PROBABLY 33 KD BETA-GALACTOSIDE-BINDING LECTIN (33 KD GBP).
ZK892.1.
CAENORHABDITIS ELEGANS.
EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
[1]
SEQUENCE FROM N.A.
STRAIN=BRISTOL N2;
LLOYD C.;
SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -I- FUNCTION: BINDS GALACTOSE (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; Z48638; G732606; -.
DR WORMPEP; ZK892.1; CE01724.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW HYPOTHETICAL PROTEIN; GALAPTIN; LECTIN; REPEAT.
FT DOMAIN 1 141 GALAPTIN 1.
FT DOMAIN 142 289 GALAPTIN 2.
SQ SEQUENCE 289 AA; 32991 MW; 3329858D CRC32;

Query Match      11.48; Score 112; DB 1; Length 289;
Best Local Similarity 29.38; Pred. No. 1.12e-03;
Matches 12; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

Db    11 INIPYRSKLTERIEPGQTLIRGKTIDESKRFINLHKDSP 51
      :::| | | | | | | | | | | | | | | | | | | | |
Qt    5 LEVPCSHALFGQLSPGVIVIRGLVLQEPKHFTVSLRDQA 45
      ::|| | | | | | | | | | | | | | | | | | | | |

RESULT          9              STANDARD; PRT; 316 AA.
ID   LEG8.RAT
AC   Q62665;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   GALECTIN-8 (30 KD S-TYPE LECTIN) (RL-30).
GN   LGALS8.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
[1]
RN   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX   MEDLINE; 95155445.
FA   HADARI Y.R., PAZ K., DEKEL R., MESTROVIC T., ACCILI D., ZICK Y.;
RJ   J. BIOL. CHEM. 270:3447-3453(1995).
CC -I- FUNCTION: POSSESS SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -I- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, CARDIAC MUSCLE, LUNG, AND BRAIN.
CC -I- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN WHOLE EMBRYOS, HIGH LEVELS IN ADULT TISSUES.
CC -I- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR   EMBL; U09824; G717032; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW GALAPTIN; LECTIN; REPEAT.
FT DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 36038 MW; 96341BC7 CRC32;

Query Match      11.18; Score 109; DB 1; Length 316;
Best Local Similarity 22.18; Pred. No. 3.21e-03;
Matches 30; Conservative 41; Mismatches 58; Indels 7; Gaps 5;

Db    181 HLSPFPEARLNASMGPRVTVVVGENVNNTATSFNDVLVAGSRDIALHLNPLRVAFVR 240
      | | | | | | | | | | | | | | | | | | | | |
Qv    4 RLEVPCKSHALPGQLSGPVIVIRGLVLQEPKHFTVSL-RDQA 62
      ||| | | | | | | | | | | | | | | | | | | |

```

```

Db 241 NSFQDAMGEERNITCPFFSGMYFEMIIYCDVREKVVAVNGVSHSLEKYKHKRFKO-LSSI 299
QY 63 IS--R--WGQKKLISAPFLFYQREFEVLVLLFQEGGLKALNG-OGLGATSMNQQALEQL 117
Db 300 DTLAVDGDIRLIDVRS 315
QY 118 RELRISGSVOLYCVHS 133

RESULT 10
ID LEG4_RAT STANDARD; PRT; 324 AA.
AC P38552;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
DE LGALS4.
GN RATTUS NORVEGICUS (RAT).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE; 93194902.
RA ODA Y., HERRMANN J., GITT M., TURCK C.W., BURLINGAME A.L.,
RA BARONDES S.H., LEFFLER H.,
RL J. BIOL. CHEM. 268:5929-5939(1993).
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FULL-LENGTH FORM IN SMALL
CC AND LARGE INTESTINE AND STOMACH BUT WAS NOT DETECTED IN OTHER
CC TISSUES INCLUDING LUNG, LIVER, KIDNEY, AND SPLEEN.
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; M73553; G294572; -.
DR PIR; A46631; A46631.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 2.
KW GALAPTIN; LECTIN; REPEAT.
FT DOMAIN 1 152 GALAPTIN 1.
FT LINKER.
FT DOMAIN 153 177
FT DOMAIN 178 324 GALAPTIN 2.
FT BINDING 257 263 BETA-GALACTOSIDE (BY SIMILARITY).
FT SEQUENCE 324 AA; 36347 MW; 9870240B CRC32;

Query Match 11.1%; Score 109; DB 1; Length 324;
Best Local Similarity 23.7%; Pred. No. 3.21e-03;
Matches 31; Conservative 38; Mismatches 56; Indels 6; Gaps 5;

Db 194 VPHYGTLOGGLTARTIIKGYVLPATKNNLIINFKVGTGDTAFHNPRIQCVRNSTM 253
QY 7 VPCSHALPQGLSPGVIIIVRGLVQEPKHFVSLR-DQAAHAPVTLIRASFADRTL--AWI 63
Db 254 NGSGSE--RKIPYNPFCAGOFFDLISRCGTDREKVFANGQHLDFSHRFQAFQVDMLE 312
QY 64 S-RMGQKKLISAPF-LFYQREFEVLVLLFQEGGLKALNGOGLGATSMNQQALEQLR 121

Db 313 IKGDITLSYVQ 323
QY 122 ISGSVOLYCVH 132

RESULT 11
ID GPT_YEAST STANDARD; PRT; 448 AA.
AC P07286;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE UDP-N-ACETYLGLUCOSAMINE--DOLICHYL-PHOSPHATE N-
DE ACETYLGLUCOSAMINEPHOSPHOTRANSFERASE (EC 2.7.8.15) (GPT) (G1PT)

```

```

DE (N-ACETYLGLUCOSAMINE-1-PHOSPHATE TRANSFERASE) (GLCNAC-1-P
DE TRANSFERASE) (TUNICAMYCIN RESISTANCE PROTEIN 1).
GN ALG7 OR TURI OR YER243C OR YBR1628.
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87203393.
RA HARTOG K.O., BISHOP B.;
RL NUCLEIC ACIDS RES. 15:3627-3627(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA ALJINOVIC G., POHL F.M., POHL T.M.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CATALYZES THE INITIAL STEP IN THE SYNTHESIS OF DOLICHOL-
CC P-P-OLIGOSACCHARIDES.
CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + DOLICHYL
CC PHOSPHATE -> UMP + N-ACETYL-D-GLUCOSAMINYL-DIPHOSPHODOLICHOL.
CC -!- ENZYME REGULATION: INHIBITED BY TUNICAMYCIN.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: SIGNIFICANT, TO OTHER SPECIES G1PT.
CC -!- SIMILARITY: SOME TO YEAST ALG1, DPM1, AND SEC19 AND TO
CC BACTERIAL MRAY.
DR EMBL; Y00126; G4685; -.
DR EMBL; Z36112; G536653; -.
DR PIR; A27522; A27522.
DR SGD; L0000078; ALG7.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; GLYCOPROTEIN;
KW ENDOPLASMIC RETICULUM.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
FT TRANSMEM 387 407 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
FT SITE 81 93 DOLICHOL RECOGNITION (POTENTIAL).
FT SEQUENCE 448 AA; 50368 MW; EFC9A131 CRC32;

Query Match 10.8%; Score 106; DB 1; Length 448;
Best Local Similarity 24.7%; Pred. No. 9.03e-03;
Matches 19; Conservative 21; Mismatches 32; Indels 5; Gaps 5;

Db 224 GLEVGCIVLAILALNDLLYFSMGPLATRDSHRFSAVLIIFLGVSLALWKNRPATV 283
QY 16 GLSPGVIIIVRGL-VIQEPKHFVTS--LRDQAAHA-PVTLIRASFADRTL-A-WI-SRWGQKK 70
Db 284 FVGDTYCYFAGMVFVAV 300
QY 71 LISAPFLFYQREFEVL 87

RESULT 12
ID LEG4_CHICK STANDARD; PRT; 134 AA.
AC P07583;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE BETA-GALACTOSIDE-BINDING LECTIN (14 KD LECTIN) (C-14).
OS GALLUS GALLUS (CHICKEN)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86130505.
RA OHYAMA Y., HIRABAYASHI J., ODA Y., OHNO S., KAWASAKI H., SUZUKI K.,

```

```

Best Local Similarity 33.3%; Pred. No. 1.30e-01;
Matches 21; Conservative 15; Mismatches 20; Indels 7; Gaps 7;

Db 182 ILCSHVLPGVPAHINRAAILAKG-CLQAVGSLs-QLRAEAG-LPVRIRASGISERD-SWL 237
      ||||| || : : : || : || : || : || : || : || : || : || : || : || :
Qy 7 VPCSHALPQGLSP-GQV-IIVRGLVLPKPHFTVSLRDQAHAHAPVILRAS-PADRTLAWI 63

Db 238 QRW 240
      ||
Qy 64 SRW 66

RESULT 14
ID YRNA_CAEEL STANDARD; PRT; 184 AA.
AC Q09610;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHELICAL 21.0 KD PROTEIN R07B1.10 IN CHROMOSOME X.
GN R07B1.10.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA KERSHAW J.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL: Z48621; G728535; --
DR WORMPEP; R07B1.10; CE01627.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
DR KW HYPOTHELICAL PROTEIN; GALAPTIN.
FT DOMAIN 1 138 GALAPTIN.
SQ SEQUENCE 184 AA; 120991 MW; EAF0FA73 CRC32;

```

```

Query Match          9.9%; Score 97; DB 1; Length 184;
Best Local Similarity 37.3%; Pred. No. 1.80e-01;
Matches 19; Conservative 11; Mismatches 17; Indels 4; Gaps 4;

Db      12  DVP$AHAI$REOL$RAG$SEI$HVR$GHV$TH$KHK$DF$VELLS-$GPHI-$V-$LHVNF 59
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      6  EVPC$HAL$POGL$SPG$VIL$VRLQ-$EPK$HTV$SLR$DQ$AAH$APVTL$RAS$F 55
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      15
ID PFEA_PSEAE STANDARD; PRT; 746 AA.
AC Q05098;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE FERRIC ENTEROBACTIN RECEPTOR PRECURSOR.
GN PFEA.
OS PSEUDOMONAS AERUGINOSA.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC PSEUDOMONADACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K407;
RX MEDLINE; 93123148.
RA DEAN C.R., POOLE K.;
RL J. BACTERIOL. 175:317-324(1993).
CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
CC ENTEROBACTIN.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- INDUCTION: REGULATED BOTH BY IRON AND ENTEROBACTIN.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC PIR; M98033; G151435; -.
DR EMBL; A40636;
DR PDB; 1A0636;
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW OUTER MEMBRANE; IRON TRANSPORT; TRANSPORT; TONB BOX; SIGNAL; RECEPTOR.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR
FT FT

```

FT SIMILAR 39 44 TONB BOX.  
FT SIMILAR 729 746 TONB C-TERMINAL BOX.  
SQ SEQUENCE 746 AA; 80967 MW; 3BC231B9 CRC32;

Query Match 9.7%; Score 95; DB 1; Length 746;  
Best Local Similarity 37.8%; Pred. No. 3.42e-01;  
Matches 14; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

Db 363 YTATLRDLTAHGEVNLPHLHGYEQTLTLGSEWTEOKL 399  
QY 36 FTVSLRDQAAHAPVTLRASFA-DRTLAWISRWGOKKL 71

Search completed: Tue Dec 15 15:55:27 1998  
Job time : 16 secs.

\*\*\*\*\*

WQREH  
(TW)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:30:42 1998; MasPar time 1134.44 Seconds  
Tabular output not generated. 1365.449 Million cell updates/sec

Title: >US-09-109-864-1  
Description: (1-865) from US09109864.seq  
Perfect Score: 865  
N.A. Sequence: 1 ttgtgtgagggcagcagaga.....aaaaaaaaaaaaaaaaaaaa 865  
Comp: aaacacccctcgtctct.....tttttttttttttttttt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est55

Database: genbank-est107  
1:em-est1 2:em\_gss1 3:em\_gss2 4:em\_gss3  
5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13  
10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17  
14:gb-est18 15:gb-est19 16:gb-est20  
17:gb-est21 18:gb-est22 19:gb-est23 20:gb-est24 21:gb-est25 22:gb-est26  
23:gb-est27 24:gb-est28 25:gb-est29 26:gb-est30 27:gb-est31  
28:gb-est32 29:gb-est33

Statistics: Mean 12.059; Variance 7.331; scale 1.645

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	463	53.5	496	24	AA311108 EST181872 Jurkat T-cell	2.06e-287
2	336	38.8	411	16	H30148 yo59h05.r1 Homo sapien	2.64e-199
3	319	36.9	422	16	H27468 yo59h05.s1 Homo sapien	1.39e-187
4	221	25.5	227	25	AA349100 ES1555685 Infant adrena	1.44e-120
5	120	13.9	660	22	AA020729 ze63a11.r1 Soares reti	2.17e-53
6	80	9.2	434	22	AA001029 ze46b12.r1 Soares reti	4.14e-28
7	61	7.1	252	12	AA754459 97SN1787 Rice Immature	9.89e-17
8	55	6.4	252	12	AA754459 97SN1787 Rice Immature	2.79e-13
9	51	5.9	448	5	R37188 yh96a12.r1 Homo sapien	4.94e-11
10	50	5.8	341	12	C84867 Dictyostelium discoide	1.77e-10
11	50	5.8	492	17	A1128239 qc33f09.x1 Soares_preg	1.77e-10
12	49	5.7	247	12	AA754458 97SN1784 Rice Immature	6.32e-10
13	49	5.7	558	7	C24325 Dictyostelium discoide	6.32e-10

14	48	5.5	151	14	C90955 Dictyostelium discoide	2.23e-09
15	48	5.5	241	14	C90027 Dictyostelium discoide	2.23e-09
16	48	5.5	252	14	C89779 Dictyostelium discoide	2.23e-09
17	48	5.5	359	14	C91502 Dictyostelium discoide	2.23e-09
18	48	5.5	578	15	C94202 Dictyostelium discoide	2.23e-09
19	48	5.5	785	26	B21585 T28D23-Sp6 TAMU Arabid	2.23e-09
20	47	5.4	125	22	N97843 2135C3 czapFDA2.1. De	7.83e-09
21	47	5.4	166	11	C84166 Dictyostelium discoide	7.83e-09
22	47	5.4	184	14	C92535 Dictyostelium discoide	7.83e-09
23	47	5.4	241	7	C24389 Dictyostelium discoide	7.83e-09
24	47	5.4	259	23	AA137934 mw01a08.r1 Soares mous	7.83e-09
25	47	5.4	267	23	AA231371 mw30a07.r1 Soares mous	7.83e-09
26	47	5.4	268	14	C91192 Dictyostelium discoide	7.83e-09
27	47	5.4	302	9	AA607600 vo48e06.r1 Barstead mo	7.83e-09
28	47	5.4	320	18	A1154582 Dictyostelium discoide	7.83e-09
29	47	5.4	339	7	C23862 Dictyostelium discoide	7.83e-09
30	47	5.4	396	14	AA980543 ua42e08.r1 Soares mous	7.83e-09
31	47	5.4	405	18	A1180802 ub76e02.r1 Soares mous	7.83e-09
32	47	5.4	476	12	AA747432 nx88f03.s1 NCI CGAP GC	7.83e-09
33	47	5.4	580	14	C91340 Dictyostelium discoide	7.83e-09
34	47	5.4	614	14	C93036 Dictyostelium discoide	7.83e-09
35	46	5.3	183	14	C93036 Dictyostelium discoide	7.83e-09
36	46	5.3	200	15	C93770 Dictyostelium discoide	7.83e-09
37	46	5.3	224	11	C84161 Dictyostelium discoide	7.83e-09
38	46	5.3	247	12	AA754458 97SN1784 Rice Immature	2.72e-08
39	46	5.3	260	15	A1049844 an21e05.x1 Gessler Wil	2.72e-08
40	46	5.3	341	12	AA748451 os56e01.s1 NCI CGAP GC	2.72e-08
41	46	5.3	378	18	A1146542 qb45a12.x1 NCI CGAP Br	2.72e-08
42	46	5.3	437	15	C94484 Dictyostelium discoide	2.72e-08
43	46	5.3	450	14	C89917 Dictyostelium discoide	2.72e-08
44	46	5.3	459	24	AA266894 mz98h12.r1 Soares mous	2.72e-08
45	46	5.3	821	26	B21581 T25117-Sp6 TAMU Arabid	2.72e-08

ALIGNMENTS

RESULT	1	AA311108	496 bp	mRNA	EST	19-APR-1997
LOCUS		EST181872	Jurkat T-cells	V Homo sapiens	CDNA 5'	end, mRNA sequence.
DEFINITION		AA311108				
ACCESSION		AA311108				
NID		Q1963435				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1	(bases 1 to 496)			
AUTHORS		Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgeral,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodak,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissen,P.S., Olsen,H., Raymond,L., Wei,X.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Fields,C., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.				
TITLE		Initial assessment of human gene diversity and expression patterns				
JOURNAL		Nature 377 (6547 Suppl), 3-174 (1995)				
RESUME		Based upon 81 million nucleotides of cDNA sequence				
COMMENT		Other ESTs: THC124424				
		Contact: Kerlavage, AR				
		Bioinformatics				
		The Institute for Genomic Research				

9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlavetigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse

# FEATURES

Location/Qualifiers

source  
1..496  
/organism="Homo sapiens"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
/db\_xref="ATCC (inhost):157086"  
/db\_xref="taxon:9606"  
/clone\_lib="Jurkat T-cells V"  
/cell\_type="T-lymphocyte"  
<1..496  
151 g 103 t 1 others

mRNA  
BASE COUNT  
ORIGIN

Query Match 53.5%; Score 463; DB 24; Length 496;  
Best Local Similarity 99.2%; Pred. No. 2.06e-287;  
Matches 482; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Db 5 TTTGTGGAGGGCAGCAGAGATACCCAGCTGGACATCTCTCTGTGTGATGAGCCCCAGG 64  
Qy 1 tttgtggaggcagcagagatgccagctggacatcttctctgtgtgagcccccagg 60  
Db 65 CTGAGGATGCTCTGCTCATGCTCTTCCCGAGGCTCTCGCTCGGGCAGGTCATCAT 124  
Qy 61 ctggagggtgcccgtgctcatgctcttcccagggtctcgtcgtggcaggtcatcata 120  
Db 125 GTACGGGACTGGTCTGTCAAGAGCCGAAAGCATTTTACTGTGAGCTGAGGGACCAAGCT 184  
Qy 121 gtacggggaactggttgcagagccgagcattttactgtgagcctgagggaccaggct 180  
Db 185 GCCCATGCTCTGTGACACTCAGGGCCTCTTGCAGACAGAACTCTGCGCTGGTTCTCC 244  
Qy 181 gccatgctcctgtgacactcagggtcctcttcgacagacagaaactctggcctggtatctcc 240  
Db 245 CGCTGGGGCAGAAAGAACTGATCTCAGCCCTTCTCTTTACCCCCAGAGATCTTT 304  
Qy 241 cgcctggggcagaagaactgatctcagccccctctcttttacccccagagatcttt 300  
Db 305 GAGGTGCTGCTCTGTTCAGAGGGAGGCTGAAGCTGGCGCTCAATGGCGAGGGCTG 364  
Qy 301 gaggtgctgctcctgtcttcagagggagggctggaagctgagcctcaatgggagggctg 360  
Db 365 GGGGCCACCATGATGAACACAGCAGCCCTGGAGCAGCTGCGGAGCTCGGGATCAGTGA 424  
Qy 361 ggggcccaccagcatgaaccagcagccctggagcagctcgggagctccgagatcagtgga 420  
Db 425 AGTGTCAGCTCTACTGTGTCACCTCTGA-GGATGGTTCCAGGGAAATACCCNACAGAA 483  
Qy 421 agtgtccagctactgtgtccactcctgaaggatgtttccagg-aaataccg-cagaaa 478  
Db 484 ACAAGA 489  
Qy 479 acaaga 484

RESULT 2 H30148 411 bp mRNA EST 16-AUG-1995  
LOCUS Y059h05.r1 Homo sapiens cDNA clone 182265 5'  
ACCESSION H30148  
NID g901058  
KEYWORDS EST.  
SOURCE human clone=182265 library=Soares breast 3NHBst vector=pH7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13RPI Rsite1=Not I Rsite2=Eco RI Adult human. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

### AUTHORS

1 (bases 1 to 411)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maria,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

### TITLE

The WashU-Merck EST Project

### JOURNAL

Unpublished (1995)

### COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 264

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

## FEATURES

source

Location/Qualifiers

1..411

/organism="Homo sapiens"

/clone="182265"

BASE COUNT 90 a 113 c 121 g 79 t 8 others

## ORIGIN

Query Match 38.8%; Score 336; DB 16; Length 411;

Best Local Similarity 95.7%; Pred. No. 2.64e-199;

Matches 360; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

Db 1 CAGGGTCTCTNACAGGCGAGGTATCATAGTACGGGGANTGGTCTTNCAGAGCCGAG 60  
Qy 91 cagggtctctcgctggcaggtcatcatagtcaggggactggtcttgcagagccgaag 150  
Db 61 CATTTACTGTGAGCTGAGGACCGAGCTNCCATCTCTGTGACACTCAGGCGCTCC 120  
Qy 151 catttactgtgagctgagggaccaggtgcccactgctctctgacactcagggcctcc 210  
Db 121 TTCGACAGAGAACTCTGGCTGGATCTCCGCTGGGGCAGAGAACTGATCTCAGCC 180  
Qy 211 ttgcagacagaactctggctggtatctccgctggggcagagaagaactgatctcagcc 270  
Db 181 CCCTTCTCTTTTACCCNCAGAGATCTTTGAGTGTCTCTCTGTTTCCAGGAGGAGG 240  
Qy 271 ccttctcttttacccccagagatttttggagtgtctctctgttccagggaggagg 330  
Db 241 CTGAAGCTGGCTCAATGGGAGGGGCTGGGGCCACCAAGCAATGAACAGCAGGCC 300  
Qy 331 ctgaagctggcgtcaatggcaggggctgggggccaacca-gca-tgaacacagagggccc 388  
Db 301 TGGAGCAGCTGCGGNGACTCCGGATCAGTGAAGTATCCAGCTTACTGTGTCTCCT 360  
Qy 389 tggagcagctgcgggagctccggatcagtggaagtgtccagctctactgtgtccactcct 448  
Db 361 NA-GGATGGTTCAGG 375  
Qy 449 gaaggtggttccagg 464

RESULT 3 H27468 422 bp mRNA EST 16-AUG-1995  
LOCUS Y059h05.s1 Homo sapiens cDNA clone 182265 3'  
DEFINITION

Db 359 TCCGCACATCTGCCAGGCGCTGTGGTTCANCTGTGGGCCCCCAGNCCCTNCCCAT 418  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 404 tccgcagctgctccagg -cctgctgttcattctgttg-ccccacgccctggccat 347  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 419 T 419  
I

Cp 346 t 346

RESULT 4  
LOCUS AA349100 227 bp mRNA EST 21-APR-1997  
DEFINITION EST55685 Infant adrenal gland II Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA349100  
NID G2001358  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;  
Homo.

REFERENCE 1 (bases 1 to 227)  
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,D.M., Fritcham,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodak,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
Kellay,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
Small,K.V., Spriggs,I.A., Utterback,T.R., Weidman,J.F., Li,Y.,  
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,  
Raymond,L.B., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE 96026280

COMMENT Other\_ESTS: THCL24424  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers  
source  
1..227  
/organism="Homo sapiens"  
/note="ECori: adrenal gland; Vector: pBluescript SK-";  
Site\_1: EcoRI; Site\_2: XhoI  
/db\_xref="ATCC (inhost):150218"  
/db\_xref="taxon:9606"  
/clone\_lib="Infant adrenal gland II"  
/sex="male"  
/dev\_stage="infant, 16 wks"  
<1..>227

mRNA  
BASE COUNT 42 a 72 c 60 g 51 t 2 others  
ORIGIN

Query Match 25.5%; Score 221; DB 25; Length 227;  
Best Local Similarity 98.2%; Pred. No. 1.44e-120;  
Matches 223; Conservative 0; Mismatches 14; Indels 0 Gaps 0;

```

Db 1  AGAGTACCAGCTGGACATCTTCTGCTGATGAGCCCGAGGCTGGAGTGCCCTGCTC 60
      |||
QY 18  agagtaccagctgacatcttctctgctgatgagcccgagctggagtgccctgctc 77
      |||
Db 61  ACATGCTTTCCTCCAGGCTCTCCGCTGGCAGGCTCATAGTACGGGACTGTGCTT 120
      |||
QY 78  acatgcttctccagggctctcctcggcgaggctcatcatagtagcgggagctgtctt 137
      |||
Db 121  GCAAGAGCGGAGCAATTTACTGTAAAGCTGAGGGACGAGGCTNCCCATGCTCTCTAAC 180
      |||
QY 138  gcaagagcggaagcatcttactgtgagcctgagggaccagggctgcccctgctctgtgac 197
      |||
Db 181  ACTCAGGCGCTCTCTTCGAGAGAGCAACTNTGGCCTGGATCTCCCGCT 227
      |||
QY 198  actcagggcctctcttcgagagaaactctggcctggatctcccgct 244
      |||

RESULT 5  AA020729 660 bp mRNA EST 30-JAN-1997
LOCUS ze63all.r1 Soares retina N2b4HR Homo sapiens cDNA clone 363644 5'
DEFINITION similar to contains LI.t1 LI repetitive element ;, mRNA sequence.
ACCESSION AA020729
NID 91484512
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
REFERENCE 1 (bases 1 to 660)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)
TITLE The WashU-Merck EST Project
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2180 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 444.
Location/Qualifiers
1. .660
/organism="Homo sapiens"
/notes="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="363644"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
<1. .5660
198 a 147 c 168 g 142 t 5 others

mrna
BASE COUNT

```

```

ORIGIN
Query Match 13.9%; Score 120; DB 22; Length 660;
Best Local Similarity 96.2%; Pred. No. 2.17e-53;
Matches 150; Conservative 0; Mismatches 1; Indels 5; Gaps 5;

Db 1  CTGAAGCTGGCGCTCAATGGCGAGGGCTGGGGGC-ACCAGCATGAACACGAGGCCCTG 59
      |||
QY 331  ctgaagctggcgctcaatggcgagggctggggccaccagcatgaaccagggccctg 390
      |||
Db 60  GAGCAGCTGGCGNGA-CTCCGGATCAGTGGAAAGTGTCCAGCTCTACTGTGTCCACTCTCTGA 118
      |||
QY 391  gagcagctggcgagctccggatcagtggaagtgtccagctctactgtgtccactctcta 450
      |||
Db 119  -GGATGTTCCAGGAAATACCGCCAGAAACAAGA 153
      |||
QY 451  aggatgttccagg-aaataccgc-agaaaacaaga 484
      |||

RESULT 6  AA0101029 434 bp mRNA EST 29-NOV-1996
LOCUS ze46b12.r1 Soares retina N2b4HR Homo sapiens cDNA clone 362015 5'
DEFINITION similar to contains Alu repetitive element; contains LI.t1 LI
repetitive element ;, mRNA sequence.
ACCESSION AA0101029
NID 91437096
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
REFERENCE 1 (bases 1 to 434)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE The WashU-Merck EST Project
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2264 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 345.
Location/Qualifiers
1. .434
/organism="Homo sapiens"
/notes="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="362015"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"

```

```

/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
<1..>434
BASE COUNT      128 a  104 c  104 g  94 t      4 others
ORIGIN

Query Match      9.2%; Score 80; DB 22; Length 434;
Best Local Similarity 94.5%; Pred. No. 4.14e-28;
Matches 104; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Db 1 ACCAGCAGCCTGAGGAGCTGCNGA-CTCCGATCAGTGGAGTNTCCAGCTCTACT 59
|||||
QY 377 accagcagcctggagcagctgggagctcggatcagtggaagtgtccagctctact 436
|||||

Db 60 GTGTCCACTCTGA-GGATGTTCCAGGGAATACCGCAGAAACAAGA 108
|||||
QY 437 gtgtccactctgaaggatgtgtccagg-aaataccgc-agaaaaacaaga 484
|||||

RESULT 7 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
DEFINITION cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa
Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@mioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT      5 a  21 c  12 g  35 t  179 others
ORIGIN

Query Match      7.1%; Score 61; DB 12; Length 252;
Best Local Similarity 14.3%; Pred. No. 9.89e-17;
Matches 32; Conservative 110; Mismatches 79; Indels 3; Gaps 3;

Db 13 CCCBAWMTTSYBCHNBVWCVASGNTMVSVN-CTBRTGTCDCCKNVNNTWTGTVNWB 71
|||||
QY 534 cctgaacaccagcatcagggcctggtccacctcgtgggtccagcagagctcag 593
|||||
Db 72 NVSGDWHYWBVNTKVDVGNHTRC-SRWRBVTMAHYDHTNCBBYNNNDYHWHBMYB 130
|||||

QY 594 gagctttg-ggctagggaagacacagagtgcaaggttctctcgaactgcaccttc 652
|||||
Db 131 BTGCTCTWCBWHTYKCTASGHTSTNYDVKSSTNTWGTBYSYDKSHGWCBBVY 190
|||||
QY 653 ctccaccaggagcctggatagctccatctgcttcaggcctgagctgactcacag 712
|||||
Db 191 HTKVSSTRATRSYTCVRKYCVMMTKVKVKYHVBBGCHBTD 234
|||||
QY 713 aggcaagtgttagactaacaaagatactccaaaatacaatgg 756
|||||

RESULT 8 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
DEFINITION cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa
Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@mioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT      5 a  21 c  12 g  35 t  179 others
ORIGIN

Query Match      6.4%; Score 55; DB 12; Length 252;
Best Local Similarity 10.0%; Pred. No. 2.79e-13;
Matches 21; Conservative 109; Mismatches 78; Indels 2; Gaps 2;

Db 41 MSVHNTCTBRGTHCDCKNVNNTWTGTVNWBVNSGVSWHYBNTKVDVGNHTRC 100
|||||
Cp 771 accacattttaagcattgtatttggagctattctgttagctcacacacttgc-t 713
|||||
Db 101 VTRMAHYDHTNCBBYNNNDYHWHBMYBTTGCTCTMWCWBHYNTKCTASGHTSTNY 160
|||||
Cp 712 cctgagtgctccagccctgaaggcagatggagccatcccgagctcctggtgag 653
|||||
Db 161 DVKSSTNTWGV-TBSYDKSMHGYWCBVYKTKVSTTRATRSYTCVRKYCVMMTKVKV 219
|||||
Cp 652 gaaggtgcagagtgtaggaacacttgcactcttgccttcctcagcagcccaagctcc 593
|||||

```

RESULT	10
LOCUS	C84867 341 bp mRNA EST 10-FEB-1998
DEFINITION	Dictyostellium discoideum slug cDNA, clone SSG869, mRNA sequence.
ACCESSION	C84867
NID	92865014
KEYWORDS	EST; EST(expressed sequence tag).
SOURCE	Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA, clone lib:SS.
ORGANISM	Dictyostellium discoideum

```

/organism="Homo sapiens"
/site="Organ: uterus; Vector: pT7T3-pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGGAGATTCGCGCGCCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
and Eco RI sites of the modified pT7T3 vector Library

```

[illegible]

ORGANISM Dictyostelium discoideum  
Eukaryota; Dictyosteliida; Dictyostelium.  
REFERENCE 1 (sites)  
AUTHORS Yoshida,M.  
TITLE Developmental cDNA in Dictyostelium discoideum  
JOURNAL Published Only in Database (1998) In press  
REFERENCE 2 (bases 1 to 151)  
AUTHORS Yoshida,M.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAR-1998) to the DBJ/EMBL/GenBank databases.  
Motonobu Yoshida, Kinki University, Research Institute of Food  
Science; Nakamachi 3327, Nara 631, Japan  
(E-mail:yoshida@ews06.nara.kindai.ac.jp,  
Tel: +81-742-43-1511(ex.3005), Fax: +81-742-43-1155)  
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.  
COMMENT Location/Qualifiers  
FEATURES  
source  
1. .151  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_lib="SS"  
/dev\_stage="slug"  
BASE COUNT 99 a 1 c 5 g 44 t 2 others  
ORIGIN

Query Match 5.5%; Score 48; DB 14; Length 151;  
Best Local Similarity 77.3%; Pred. No. 2.23e-09;  
Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 34 TTATTATTATTTTGTTCATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 93  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 778 ttcttattattattattgtggtcaataataataaggttatttatttataaaaaa 837

Db 94 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 121  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 838 aaaaaaaaaaaaaaaaaaaaaa 865

RESULT 15  
LOCUS C90027 241 bp mRNA EST 15-APR-1998  
DEFINITION Dictyostelium discoideum slug cDNA, clone SSG112, mRNA sequence.  
ACCESSION C90027  
NID 93059647  
KEYWORDS EST; EST(expressed sequence tag).  
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,  
clone\_lib:SS.  
ORGANISM Dictyostelium discoideum  
Eukaryota; Dictyosteliida; Dictyostelium.  
REFERENCE 1 (bases 1 to 241)  
AUTHORS Urushihara,H.  
TITLE Developmental cDNA in Dictyostelium discoideum  
JOURNAL Published Only in Database (1998) In press  
REFERENCE 2 (bases 1 to 241)  
AUTHORS Urushihara,H.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAR-1998) to the DBJ/EMBL/GenBank databases. Hideko  
Urushihara, University of Tsukuba, Institute of Biological  
Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan  
(E-mail:gxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,  
Fax:+81-0298-53-6614)  
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.  
COMMENT Location/Qualifiers  
FEATURES  
source  
1. .241  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_lib="SS"  
/dev\_stage="slug"  
BASE COUNT 153 a 15 c 9 g 62 t 2 others  
ORIGIN

Query Match 5.5%; Score 48; DB 14; Length 241;  
Best Local Similarity 90.0%; Pred. No. 2.23e-09;

Matches 54; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 159 AAAAAAAAAACAATTTTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 218  
|| ||| ||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Qy 806 aataataataaggttatttataaaaaataaaaaataaaaaataaaaaa 865

Search completed: Tue Dec 15 18:57:34 1998  
Job time : 1612 secs.

\*\*\*\*\*

W P E R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:00:04 1998; MasPar time 1259.73 Seconds

Tabular output not generated. 1513.252 Million cell updates/sec

Title: >US-09-109-864-1  
Description: (1-865) from US09109864.seq  
Perfect Score: 865  
N.A. Sequence: 1 ttgttgaggcagcagaga.....aaaaaaaaaaaaaaaaaaaa 865  
Comp: aaacacccctcgctctct.....tttttttttttttttttt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb155

1:em\_ba 2:em\_fun 3:em\_htg 4:em\_hum1 5:em\_hum2 6:em\_in  
7:em\_om 8:em\_or 9:em\_ov 10:em\_pat 11:em\_ph 12:em\_pl  
13:em\_ro 14:em\_vl

Database: genbank107

15:gb\_bal 16:gb\_ba2 17:gb\_htg 18:gb\_in 19:gb\_om 20:gb\_ov  
21:gb\_pat 22:gb\_ph 23:gb\_pl1 24:gb\_pl2 25:gb\_pr1  
26:gb\_pr2 27:gb\_pr3 28:gb\_ro 29:gb\_st 30:gb\_sts 31:gb\_sy  
32:gb\_un 33:gb\_vl

Statistics: Mean 11.899; Variance 15.521; scale 0.767

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
c 1	477	55.1	35414	26	Human Chromosome 11 Co	2.20e-137
2	54	6.2	2982	23	Dictyostelium discoideu	3.81e-04
3	48	5.5	1927	25	H.sapiens mRNA for 50	1.50e-02
4	48	5.5	2787	18	P.falciparum gene for	1.50e-02
5	48	5.5	24549	17	*** SEQUENCING IN PROG	1.50e-02
6	47	5.4	144	21	Sequence 26 from patent	2.74e-02
7	47	5.4	766	23	Sequence 1 from patent	2.74e-02
8	47	5.4	2606	23	Dictyostelium discoide	2.74e-02
9	47	5.4	10772	18	Drosophila melanogaste	2.74e-02
10	46	5.3	460	18	Lucilia cuprina cuticl	4.97e-02
11	46	5.3	1744	23	A.thaliana mRNA for th	4.97e-02
12	46	5.3	1999	28	Mus musculus putative	4.97e-02
13	46	5.3	3795	19	Bos taurus Fanconi ane	4.97e-02

14	46	5.3	10772	18	AF012089	Drosophila melanogaste	4.97e-02
15	45	5.2	472	33	HCU45456	Hog cholera virus stra	9.01e-02
16	45	5.2	542	18	PFNRSUB	P.falciparum non-repet	9.01e-02
17	45	5.2	553	18	HAY12278	H.armigera mRNA for pu	9.01e-02
18	45	5.2	1438	18	DMU86867	Drosophila melanogaste	9.01e-02
19	45	5.2	1606	18	DDU41222	Dictyostelium discoide	9.01e-02
20	45	5.2	1690	28	RNU716551	Rattus norvegicus mucu	9.01e-02
21	45	5.2	2253	18	AF032396	Hyalophora cecropia ar	9.01e-02
22	45	5.2	2794	18	DDU48271	Dictyostelium discoide	9.01e-02
23	45	5.2	3437	21	AS2184	Sequence 17 from Paten	9.01e-02
24	45	5.2	3437	21	SOPULSPO	S.oleracea L. mRNA for	9.01e-02
25	45	5.2	4826	26	HSRANBP5	H.sapiens mRNA for Ran	9.01e-02
26	45	5.2	10003	18	PFSC04009	Plasmodium falciparum	9.01e-02
27	45	5.2	13445	18	PFSC04059	Plasmodium falciparum	9.01e-02
28	45	5.2	76848	17	PFMALIP2	Plasmodium falciparum	9.01e-02
29	45	5.2	129169	17	AC005506	*** SEQUENCING IN PROG	9.01e-02
30	44	5.1	829	26	HSAJ3352	Homo sapiens mRNA for	1.63e-01
31	44	5.1	1086	18	HAY12282	H.armigera mRNA for pu	1.63e-01
32	44	5.1	2483	23	DDIL3A	Dictyostelium discoide	1.63e-01
33	44	5.1	3363	18	CEMCEIA	C.elegans mRNA for myo	1.63e-01
34	44	5.1	7218	21	I66494	Sequence 14 from Paten	1.63e-01
35	44	5.1	8574	18	AF057019	Dictyostelium discoide	1.63e-01
36	44	5.1	151018	17	PFMALIP6	Plasmodium falciparum	1.63e-01
37	43	5.0	340	21	I22446	Sequence 27 from patent	2.92e-01
38	43	5.0	350	21	I22433	Sequence 14 from patent	2.92e-01
39	43	5.0	1013	28	RNU75932	Rattus norvegicus CAMP	2.92e-01
40	43	5.0	1136	21	I08009	Sequence 3 from Patent	2.92e-01
41	43	5.0	1558	21	P03671	DNA sequence coding fo	2.92e-01
42	43	5.0	166401			Sequence 7 from patent	2.92e-01
43	43	5.0	1718	26	HSAJ3198	Homo sapiens mRNA for	2.92e-01
44	43	5.0	2874	18	S82651	RNA polymerase II larg	2.92e-01
45	43	5.0	7218	21	I66494	Sequence 14 from patent	2.92e-01

# ALIGNMENTS

RESULT 1	U73641	35414 bp	DNA	PRI	25-JUL-1997
LOCUS	Human Chromosome 11	Cosmid cSRL187d6	complete sequence.		
DEFINITION	U73641				
ACCESSION	U73641				
NID	G2281065				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 35414)				
AUTHORS	Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D., Davies,J., Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.				
TITLE	Template				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 35414)				
AUTHORS	Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D., Davies,J., Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUL-1997)				
COMMENT	Genome Science and Technology Center, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA				
FEATURES	On Jul 25, 1997 this sequence version replaced gi:1737196.				
source	Location/Qualifiers				
	1..35414				
	/organism="Homo sapiens"				

[illegible]

```

/db_xref="taxon:44689"
/tissue_type="germinating spore"
984..1253
/gene="CelB"
/number=1
984..2881
/gene="CelB"
Join(1196..1253,1341..2881)
/gene="CelB"
/note="spore germination-specific protein; formerly called
270-11"
/codon_start=1
/product="cellulose binding protein"
/db_xref="PID:g167881"
/translocation="MKNIYSLFLFALLISATFANNAFIHWNSDSISKLTGQIGDTI
SFTSDGNHDDKSDGVSFSGSLTNPGLFKVTLKEGNIETSSVDEGLSATI
VSSGGQIPVTTSSTTSDGSSSPSTSTTSATTSAGSATTTGEPITDGSNGA
SSTGSGTGSAITTTSSSDNSDGSVGTSTSPAITSSGSIIDPSPPTDSSS
NSGGYSSSIYNGVSCLLTQTQDAFDSWTDNIYTVQVNLNTNIGTUSVESVILPT
DNMSLIHWELVYDSTLFTPYRKAGINPEETIFGYISRNSTDVTFALSPGTCSD
SSSTPTPTETPTETPTETPTETPTETPTETPTETPTETPTETPTETPTETPT
SEIETPTETPTETPTETPTETPTETPTETPTETPTETPTETPTETPTETPT
INCEVDHYIQVEATIVNQGSTRPISSNFYSDAQINSVEKNTYKLPWFSTIPVG
GSTFTGYIYKSAELSDLEGVQYIC"
1254..1340
/gene="CelB"
1341..>2881
/gene="CelB"
/number=2
BASE COUNT 1168 a 417 c 306 g 1091 t
ORIGIN
Query Match 6.2%; Score 54; DB 23; Length 2982;
Best Local Similarity 79.3%; Pred. No. 3.81e-04;
Matches 73; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 349 TTTTATTTTTTATTATTTTTTTTTTTGTTTTTAAATAATAAAAAATTTAATTAA 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 774 ttattcttatttatttatttatttggccaataataataaaggttatttatttaa 833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 409 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 834 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 3
LOCUS HSEPMG50 1927 bp RNA PRI 30-JUN-1993
DEFINITION H.sapiens mRNA for 50 kDa erythrocyte plasma membrane glycoprotein.
ACCESSION X64594 S46252
NID g31194
KEYWORDS glycoprotein; plasma glycoprotein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1927)
Direct Submission
Ridgwell,K.
Submitted (23-NAR-1992) K. Ridgwell, University of Bristol, Dept of
Biochemistry, School of Medical Sciences, University Walk, Bristol
BS8 1TD, UK
2 (bases 1 to 1927)
Ridgwell,K., Spurr,N.K., Laguda,B., MacGeoch,C., Avent,N.D. and
Tanner,M.J.
Isolation of cDNA clones for a 50 kDa glycoprotein of the human
erythrocyte membrane associated with Rh (rhesus) blood-group
antigen expression
Biochem. J. 287 (Pt 1), 223-228 (1992)
93038558
REFERENCE 2 Location/Qualifiers
AUTHORS 1. .1927
/organism="Homo sapiens"
/db_xref="taxon:9606"
JOURNAL
MEDLINE
FEATURES
source

```

```

exon      /cell_line="T9/96 blood stage culture"
/clone_lib="Lambda gt11"
/chromosome="8"
1. .806
/number=1
join(1. .806,982. .2787)
/product="STARP antigen"
join(735. .806,982. .2724)
CDS       /codon_start=1
/product="STARP antigen"
/db_xref="PID:499325"
/translation="MHIFYKTAFTLSIWTLLYSNKLKCNFYNNNNLSTYVIKH
NRFLEYSOSPLGGYSAALKLVNSKSGTNTYKNSENTNTNNIPSSSTYINTR
LAANNSTITSTKVDNNKNTIKLGNNSSTINTNSTENTATKKVTVIYNQILITG
NNININTNSDNTSNTNLNTDNTSTTKRLDNTINTQNLTTSTNTTSTONNINT
KPIDNNNDIKSTNTGTETDNKNTDKATDNNNTITTTDNTNTVISTDNSKTN
VISTDNTNTISTDNDNADTLTDNDNNTDILTDNNNTDASTDNDNADKATDNN
NNTKATDNNNTKILSPDNNNTKTSIDNNNTNTKATDNNNTKTSIDNNNTKTS
TDNNNTKTSIDNNNTKTSIDNNNTNTISTDNNNTNTQYVFANNYNETTSDELNK
DSCDYSEKENIKSMINAYLDKLDLETVRKHSIDSTCEKKNRPNRQIHLNKNM
YNIIRFIVIVIAFNWSEVYKYVGKILAFALYMLIN"
807. .981
/number=1
exon      982. .2787
/number=2
BASE COUNT 1411 a 306 c 210 g 860 t
ORIGIN

Query Match      5.5%; Score 48; DB 18; Length 2787;
Best Local Similarity 77.3%; Pred. No. 1.50e-02;
Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 444 TTCATTATTCCTTTATTTATTTAATTAACCAATAATAATATATATATAGATTAAAGAAAA 503
Qy 778 tctttattattattatttattgtgcataataataaagggtattattattataaaaaa 837

Db 504 AAAAAAATAAAAAAATAAAAAA 531
Qy 838 aaaaaaataaaaaaataaaaaa 865

RESULT 5
LOCUS      AC005308 244649 bp DNA HTG 20-AUG-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Plasmodium falciparum 3D7 chromosome
12 PFYAC492 genomic sequence; HTGS phase 1, 36 unordered pieces.
ACCESSION  AC005308
NID        93445189
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     malaria parasite P. falciparum.
ORGANISM   Plasmodium falciparum
REFERENCE  1 (bases 1 to 244649)
AUTHORS    Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE      Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
JOURNAL    Plasmodium falciparum 3D7 chromosome 12
REFERENCE  2 (bases 1 to 244649)
AUTHORS    Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE      Direct Submission
JOURNAL    Submitted (21-JUL-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT    On Aug 20, 1998 this sequence version replaced gi:3334987.
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 36 contigs for
* which the order is not known: their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.

```

FEATURES  
Location/Qualifiers  
1. .244649

```

* 1 11189: contig of 11189 bp in length
* 24361: gap of unknown length
* 37253: contig of 13172 bp in length
* 42268: gap of unknown length
* 51249: contig of 12892 bp in length
* 58344: gap of unknown length
* 56377: contig of 5015 bp in length
* 76265: gap of unknown length
* 90094: contig of 8981 bp in length
* 103795: gap of unknown length
* 113640: contig of 7095 bp in length
* 122079: gap of unknown length
* 130477: contig of 8033 bp in length
* 138590: gap of unknown length
* 146666: contig of 8077 bp in length
* 154737: gap of unknown length
* 162631: contig of 8077 bp in length
* 170496: gap of unknown length
* 177936: contig of 8077 bp in length
* 185316: gap of unknown length
* 191922: contig of 8077 bp in length
* 198084: gap of unknown length
* 203810: contig of 8077 bp in length
* 208771: gap of unknown length
* 213163: contig of 8077 bp in length
* 217459: gap of unknown length
* 221544: contig of 8077 bp in length
* 225441: gap of unknown length
* 228635: contig of 8077 bp in length
* 231186: gap of unknown length
* 233697: contig of 8077 bp in length
* 236048: gap of unknown length
* 238359: contig of 8077 bp in length
* 240643: gap of unknown length
* 242673: contig of 8077 bp in length
* 242674: gap of unknown length
* 244649: contig of 8077 bp in length.

```

[illegible]



KEYWORDS pyk10 gene; thioglucosidase; thioglucoside glucosylhydrolase.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryotes; mitochondrial eukaryotes; Viridiplantae;  
Magnoliopsida; Embryophyta group; Embryophyta; Magnoliophyta;  
Magnoliopsida; Caprales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 1744)  
AUTHORS Schmidt, K.P., Burrows, P.R., Davies, K.G., Kammerloher, W.,  
Schaeffner, A.R., Buck, F., Cal, D., and Grundler, F.M.W.  
TITLE A root specific myrosinase in Arabidopsis responding to cyst  
nematode infection  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1744)  
AUTHORS Schmidt, K.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-1995) K.P. Schmidt, Institut fuer  
Phytopathologie, Hermann-Rodewald-Str. 9, D- 24118 Kiel, FRG  
COMMENT Related sequences: AT4257 and Z34217.  
FEATURES Location/Qualifiers  
source 1..1744  
/organism="Arabidopsis thaliana"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"  
/dev\_stage="28 days old"  
/tissue\_type="root tissue"  
/clone\_lib="pCDM8 COS cell expression library"  
/clone\_full\_length\_in\_pBsc\_M13"  
6..77  
/gene="pyk 10"  
6..1580  
/gene="pyk 10"  
/EC\_number="3.2.3.1"  
/codon\_start=1  
/product="thioglucoside glucosylhydrolase"  
/db\_xref="PID:g984052"  
/translation="MWLQKPLIGLLLTIVASPNADGVCPCSNKLSRASPEGF  
LFGTAAAYQEGAINETCRGPAWDIYCRYPKCNNDNDGAVDFPHRREDIQLM  
KNLTDAFRMSIAWPRIFPHGRKEKGSQAGVQFYHDLIDELKNGITPFTVFWHT  
PODLEIYGGFLSERIVKDFREYADFVQYGGYKHWITNEPWFVHAGYDVGKKA  
PGRCSSVNAKCDGRSGYAYLVTHNLLISHAEAVEAYRCKCKGKGIGTANSPAW  
FEADLADSDGASIDRALDFILGHLDTTTFGDPYQIMKDIVGHRLPKFTTEQAKL  
KASTDFVGLAYTVSFNSHLEKPPSPRWQDSLIWESNAQNYAIGSKPLTALN  
YISGRFLKYYIKDKTANPEIMENGIGELGSDSVAVGTADHNRKRYLQRLHLS  
MQEAVCDIKVNTVGYFWSLLDNFEWODGYKNRFLGYVDFKNNLTRYEKESGKYKD  
FLSQGVRPSALKKDEL"  
6..1580  
/gene="pyk 10"  
polyA\_signal 1654..1659  
BASE COUNT 561 a 332 c 380 g 471 t  
ORIGIN  
Query Match 5.3%; Score 46; DB 23; Length 1744;  
Best Local Similarity 91.1%; Pred. No. 4,97e-02;  
Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 1654 AATAAAGGCTATTATTTTAAAAAATAAATAAATAAATAAATAAATAA 1709  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 810 aataataaggtattattttataataataataataataataataataata 865  
gene  
/gene="pyk 10"  
polyA\_signal 561 a 332 c 380 g 471 t  
BASE COUNT 561 a 332 c 380 g 471 t  
ORIGIN  
Query Match 5.3%; Score 46; DB 23; Length 1744;  
Best Local Similarity 91.1%; Pred. No. 4,97e-02;  
Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 1654 AATAAAGGCTATTATTTTAAAAAATAAATAAATAAATAAATAAATAA 1709  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 810 aataataaggtattattttataataataataataataataataataata 865  
gene  
/gene="pyk 10"  
polyA\_signal 561 a 332 c 380 g 471 t  
BASE COUNT 561 a 332 c 380 g 471 t  
ORIGIN  
Query Match 5.3%; Score 46; DB 28; Length 1999;  
Best Local Similarity 91.1%; Pred. No. 4,97e-02;  
Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 1942 AATAAAGGCTATTATTTTAAAAAATAAATAAATAAATAAATAAATAA 1997  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 810 aataataaggtattattttataataataataataataataataataata 865  
RESULT 13  
LOCUS BTU73585 3795 bp mRNA MAM 15-AUG-1997  
DEFINITION Bos taurus Fanconi anemia group C mRNA, complete cds.  
ACCESSION U73585  
NID 92326995  
KEYWORDS Bos taurus.  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Euthera; Artiodactyla; Ruminantia; Pecora;  
Bovidae; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3795)  
AUTHORS Wong, J.C.Y., Alon, N., and Buchwald, M.  
TITLE Cloning of the bovine and rat Fanconi anemia group C cDNA  
JOURNAL Mamm. Genome 8 (7), 522-525 (1997)  
MEDLINE 97341119  
REFERENCE 2 (bases 1 to 3795)  
AUTHORS Wong, J.C.Y., Buchwald, M., and Alon, N.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1996) Genetics, Hospital for Sick Children, 555  
University Ave., Toronto, Ontario M5G 1X8, Canada  
FEATURES Location/Qualifiers  
source 1..3795  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/tissue\_type="liver"  
727..2430  
/note="similar to human FAC protein, GenBank Accession  
Numbers X66893 and X66894"  
/codon\_start=1  
/product="Fanconi anemia group C protein"  
/db\_xref="PID:g2326996"

Unpublished  
2 (bases 1 to 1999)  
AUTHORS Johnson, C.A., White, D., O'Neill, L.P. and Turner, B.M.  
TITLE Direct Submission  
JOURNAL Submitted (02-DEC-1996) Department of Anatomy, University of  
Birmingham, Birmingham B15 2TT, U.K.  
COMMENT On Sep 2, 1997 this sequence version replaced gi:1737470.  
FEATURES Location/Qualifiers  
source 1..1999  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1..1999  
/gene="HD1"  
15..1463  
/gene="HD1"  
/note="similar to human histone deacetylase HD1 and yeast  
RPD3"  
/product="putative histone deacetylase"  
/codon\_start=1  
/db\_xref="PID:g2347180"  
/translation="MAOTGTRKRYCYDDYDGVNYYGQGHMPXPHRIRMTNLLN  
YGLYRMEIYRPHKANAEMTKYHSDDYIKFLRSIRPDNMSEYKMQMORFVNGDCPV  
FDGLPEFCQLSTGGSVASAKLNKQOTDIANNWAGLHAKKSEAGSCFYVNDIVLAI  
LELLYHORVLYIDIDIHGGDVEAFYTORVMTVSEHKYGEYFPGTGLRDIGAGK  
GKIYAVNPLRDGIDDESEYALFKFVMSKVMEMFOPSAVYLCQSDSLSGDRLCGFL  
TIKHAKCVFEVFKSNFLMPLMIGGGYTIIRNARCWTYETAVALDEIPEINLPYNDYF  
EYFGDFKLHISPSNNTQNTNEYLEKIQRLFNRLMPLPHAPGVQMAIPEDAIPPE  
SGDEDEDPKRISSICDKRIACEEFSDESGECCGKNSNFKKAKRVKTEDEKE  
KDPEKKEVTEEEKTKPEAKGVKEVLA"  
BASE COUNT 583 a 467 c 480 g 467 t 2 others  
ORIGIN  
Query Match 5.3%; Score 46; DB 28; Length 1999;  
Best Local Similarity 91.1%; Pred. No. 4,97e-02;  
Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Db 1942 AATAAAGGCTATTATTTTAAAAAATAAATAAATAAATAAATAAATAA 1997  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 810 aataataaggtattattttataataataataataataataataataata 865  
RESULT 13  
LOCUS BTU73585 3795 bp mRNA MAM 15-AUG-1997  
DEFINITION Bos taurus Fanconi anemia group C mRNA, complete cds.  
ACCESSION U73585  
NID 92326995  
KEYWORDS Bos taurus.  
SOURCE Bos taurus.  
ORGANISM Bos taurus.  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Euthera; Artiodactyla; Ruminantia; Pecora;  
Bovidae; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3795)  
AUTHORS Wong, J.C.Y., Alon, N., and Buchwald, M.  
TITLE Cloning of the bovine and rat Fanconi anemia group C cDNA  
JOURNAL Mamm. Genome 8 (7), 522-525 (1997)  
MEDLINE 97341119  
REFERENCE 2 (bases 1 to 3795)  
AUTHORS Wong, J.C.Y., Buchwald, M., and Alon, N.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1996) Genetics, Hospital for Sick Children, 555  
University Ave., Toronto, Ontario M5G 1X8, Canada  
FEATURES Location/Qualifiers  
source 1..3795  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/tissue\_type="liver"  
727..2430  
/note="similar to human FAC protein, GenBank Accession  
Numbers X66893 and X66894"  
/codon\_start=1  
/product="Fanconi anemia group C protein"  
/db\_xref="PID:g2326996"

	polya_site	1023 a	868 c	886 g	1018 t	
BASE COUNT						
ORIGIN						
Query Match	5.3%	Score 46;	DB 19;	Length 3795;		
Best Local Similarity	87.1%	Pred. No. 4.97e-02;				
Matches	54;	Conservative	0;	Mismatches	8;	Indels 0; Gaps 0;
Db	3732	CAATAATAAACAAGCAGACGATTCTATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	3791			
Qy	804	caataataataaggtatt				
Db	3792	AA 3793				
Qy	864	aa 865				
RESULT	14					
LOCUS	AF012089	10772 bp	DNA	INV	05-AUG-1997	
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (CP1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.					
ACCESSION	AF012089					
NID	g2305220					
KEYWORDS	fruit fly.					
SOURCE	Drosophila melanogaster					
ORGANISM	Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 4546 to 4553)					
AUTHORS	Gray,Y.H., Tanaka,M.M. and Sved,J.A.					
TITLE	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion					
JOURNAL	Genetics 144 (4), 1601-1610 (1996)					
MEDLINE	97132596					
REFERENCE	2 (bases 1 to 10772)					
AUTHORS	Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.					
TITLE	Structure of the cysteine proteinase (CP1) gene of Drosophila melanogaster and associated mutational effects					
JOURNAL	Unpublished					
REFERENCE	3 (bases 1 to 10772)					
AUTHORS	Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.					
TITLE	Direct Submission					
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia					
FEATURES	Location/Qualifiers					
source	1..10772					
mRNA	/organism="Drosophila melanogaster"					
	/db_xref="taxon:7227"					
	join(872)..1000,2310..2426,6476..6690,6751..7707)					
	/genes="Cpi"					
gene	/product="cysteine protease"					
	872..7707					
	/genes="CPI"					
exon	872..1000					
	/genes="CPI"					
	/number=1					
intron	1001..2309					
	/genes="CPI"					
exon	2310..2426					
	/genes="CPI"					
	/number=2					

REFERENCE 1 (bases 1 to 472)  
AUTHORS Mueller,H.M., Stark,R., Thiel,H.-J. and Pfaff,E.  
TITLE Comparison of different hog cholera virus strains: Implications for  
viral pathogenicity  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 472)  
AUTHORS Mueller,H.M.  
TITLE Direct Submission  
JOURNAL Submitted (11-JAN-1996) Hubert M. Mueller, Federal Research Centre  
for Virus Diseases of Animals, Paul-Ehrlich Str. 28, P.O. Box 1149,  
Tuebingen, D-72001, Germany  
FEATURES  
source  
1. .472  
/organism="Hog cholera virus"  
/strain="Riems, C"  
/note="vaccine"  
/db\_xref="taxon:11096"  
CDS  
<1. .216  
/note="putative NS5 protein coding region"  
/codon\_start=1  
/product="polyprotein"  
/db\_xref="PID:g1181845"  
/translation="NADRLVSSKGNRYIPGEGHTLQGRHYEELVLARKQINNFOGTD  
RYNLGPVNMVLRRLRVMMTLIGRGA"  
3'UTR  
BASE COUNT 145 a 85 c 103 g 139 t  
ORIGIN  
Query Match 5.2%; Score 45; DB 33; Length 472;  
Best Local Similarity 86.9%; Pred. No. 9.01e-02;  
Matches 53; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Db 277 TTTTTCCTTTTTTTTTTTTTTTTTTTTATTATTATTAGATATTATTATTATTATTATT 336  
Cp 865 ttttttttttttttttttttttttttttttttttaataaataaaccttattattatt 806  
Db 337 T 337  
Cp 805 t 805

Search completed: Tue Dec 15 18:30:24 1998  
Job time : 1820 secs.

\*\*\*\*\*

W P S R E H (TW)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nm n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:57:53 1998; MasPar time 127.57 Seconds  
Tabular output not generated. 922.536 Million cell updates/sec

Title: >US-09-109-864-1  
Description: (1-865) from US09109864.seq  
Perfect Score: 865  
N.A. Sequence: 1 tttgtggaggcgcagagaga.....aaaaaaaaaaaaaaaaaaaa 865  
Comp: aaacacctccgtcgtctct.....tttttttttttttttt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-genseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40

Statistics: Mean 9.274; Variance 8.844; scale 1.049

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	5.4	766	3 N40162	Sequence of prepropar	5.38e-07
2	46	5.3	91	9 Q51746	Oligonucleotide probe	1.28e-06
3	45	5.2	3437	21 T31478	Spinach debranching e	3.02e-06
4	44	5.1	1066	8 Q49943	Human anti-HBs light	7.13e-06
5	44	5.1	8920	11 Q62924	Carbamoyl-phosphate-s	7.13e-06
6	43	5.0	340	16 Q93933	Rat allograft inflam	1.67e-05
7	43	5.0	350	16 Q93938	Rat allograft inflam	1.67e-05
8	43	5.0	1078	4 Q25785	Protein synthesis inh	1.67e-05
9	43	5.0	1136	2 N70435	Sequence encoding ins	1.67e-05
10	43	5.0	1364	40 V15228	Heterodimeric human i	1.67e-05
11	43	5.0	1364	26 T48099	Human interleukin-12	1.67e-05
12	43	5.0	1396	4 Q24371	Natural killer cell s	1.67e-05
13	43	5.0	1558	2 Q10896	Encodes Xenopus Bone	1.67e-05

14	43	5.0	3581	10 Q44391	Sequence of murine OS	1.67e-05
15	42	4.9	259	38 V00423	3' fragment of clone	3.91e-05
16	42	4.9	515	38 V02881	Human HMGI-C aberrant	3.91e-05
17	42	4.9	756	12 Q73729	Rape abscission/dehis	3.91e-05
18	42	4.9	1558	22 T28255	Survival motor neuron	3.91e-05
19	42	4.9	1560	22 T18828	Human survival motor	3.91e-05
20	42	4.9	1582	22 T18831	Human survival motor	3.91e-05
21	42	4.9	1582	22 T28259	Survival motor neuron	3.91e-05
22	42	4.9	1710	8 Q47966	Rape acyl-ACP thioest	3.91e-05
23	42	4.9	1738	10 Q56241	Sequence encoding ant	3.91e-05
24	42	4.9	3073	3 Q14635	Clone associated with	3.91e-05
25	42	4.9	3073	21 T34371	Plasmid pATG29 (ATCC	3.91e-05
26	41	4.7	198	7 Q42784	Ligand-induced gene,	9.09e-05
27	41	4.7	259	38 T91303	Human HI075-1 secrete	9.09e-05
28	41	4.7	688	18 T16856	Integrin subunit beta	9.09e-05
29	41	4.7	867	38 V02135	Human secreted protei	9.09e-05
30	41	4.7	867	39 T88057	Partial cDNA clone en	9.09e-05
31	41	4.7	877	39 V04635	Human cytoplasmic dom	9.09e-05
32	41	4.7	1420	35 T72167	Alzheimer's disease D	9.09e-05
33	41	4.7	1441	38 T90173	Oil seed rape cystein	9.09e-05
34	41	4.7	1474	38 T90174	Oil seed rape cystein	9.09e-05
35	41	4.7	1569	1 N90711	Recombinant activatio	9.09e-05
36	41	4.7	1640	9 Q50575	Asparaginylendopeptid	9.09e-05
37	41	4.7	1910	9 Q50579	Asparaginylendopeptid	9.09e-05
38	41	4.7	2116	39 T93541	Human hyaluronan synt	9.09e-05
39	41	4.7	2589	30 T61590	Human c-IAP1.	9.09e-05
40	41	4.7	3504	1 Q05304	Clone Imd2 encoding i	9.09e-05
41	41	4.7	3850	1 N81634	Sequence of new plasm	9.09e-05
42	41	4.7	3851	33 T79857	Yellowtail tuna DNA e	9.09e-05
43	41	4.7	3933	16 T03885	Human mucosal lymphoc	9.09e-05
44	41	4.7	5108	9 Q53403	Sequence encoding all	9.09e-05
45	41	4.7	8239	29 T61981	Human ataxia and rad	9.09e-05

ALIGNMENTS

RESULT 1  
ID N40162 standard; cDNA; 766 BP.  
AC N40162;  
DE 16-FEB-1992 (first entry)  
DT Sequence of preproparathyroid cDNA.  
KW Proteolytic cleavage; trypsin; protoprotein; precursor; proteolytic;  
KW processing; ss.  
OS Bos taurus.  
FH Key Location/Qualifiers  
FT cds 111..460  
FT /\*tag= a  
PN W08401173-A.  
PD 29-MAR-1984.  
PF 09-SEP-1983; U01361.  
PR 15-SEP-1982; US-418537.  
PR 22-DEC-1982; US-452339.  
PA (IMMU-) IMMUNO NUCLEAR CORP.  
PI Mahoney WC;  
DR WPI; 84-088522/14.  
DR P-PSDB; P40209.  
PT Prodn. of protein in yeast transformed to express precursor - by  
PT use of enzyme(s) naturally present in yeast  
PS Example; Fig 2; 28pp; English.  
CC The inventors claim plasmid pYEM-1 and yeast transformed by a plasmid  
CC pYEM-1. The yeast is pref. Saccharomyces cerevisiae or S. pombe.  
CC Yeast strain X1069-2D transformed by plasmid pYEM-1 is deposited as  
CC NRRL Y-15153. pYEM-1 is constructed by the ligation of a DNA  
CC fragment (obtd. in stages from a cDNA sequence coding for bovine  
CC preproparathyroid hormone) into the Hind III site of modified  
CC plasmid yep-13.  
SQ Sequence 766 BP; 262 A; 142 C; 158 G; 204 T;

Query Match 5.4%; Score 47; DB 3; Length 766;  
Best Local Similarity 91.2%; Pred. No. 5.38e-07;  
Matches 52; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 664 aaataaatgaagttaattataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 720

PT	DNA encoding a de-branching enzyme - for modifying the branching
DT	degree of amylopectin starch synthesised in plants
PT	Claim 1; page 64-67; 90pp; English.
CC	A partial cDNA clone (r31478) codes for a spinach debranching enzyme
CC	(DE) (198371) useful for modifying the branching degree of amylopectin
CC	synthesised in plants. It was obt'd. by PCR amplification of cDNA
CC	from a spinach leaf cDNA library using probes (r31480-81) based on
CC	DE peptides (198385-86). The clone can be incorporated into a vector
CC	and used for prodn. of recombinant DE, or used in the breeding of
CC	transgenic plants; transgenic potatoes showed increased DE activity
CC	and produced a starch having altered viscosity and gel stability
CC	properties. Ribozyme or antisense sequences can be used to reduce
CC	the DE activity of transgenic plants. The modified starches have
CC	food and industrial applns.
CC	Sequence 3437 BP; 1002 A; 676 C; 727 G; 1032 T;
CC	Sequence 3437 BP; 1002 A; 676 C; 727 G; 1032 T;

PN	WO9412643-A.
PD	09-JUN-1994.
PF	02-DEC-1993.
PP	03-DEC-1992.
PR	16-DEC-1992.
PR	(UNIX ) UNISEARCH LTD.
PA	PI Flores NV, Osullivan WJ, Stewart TS;
PI	WPI; 94-200271/24.
DR	P-P8DB; R55694.
DR	PT Nucleic acid encoding carbamoyl phospho

PT isolated from Plasmodium falciparum, used to develop prods. for  
PT the treatment of malaria.  
PS Disclosure; Page 6-16; 31pp; English.

CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II  
CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes  
CC a protein that includes 2 insert sequences not found in other CPSII  
CC proteins. The first separates the putative structural subdomain and  
CC the glutaminase subdomain of the glutamine-amidotransferase subunit  
CC of CPSII, while the second separates 2 ATP binding subdomains of the  
CC CPSII subunit, CP5a and CP5b.  
SQ Sequence 8920 BP; 3836 A; 774 C; 1232 G; 3078 T;

Query Match 5.18; Score 44; DB 11; Length 8920;  
Best Local Similarity 76.3%; Pred. No. 7.13e-06;  
Matches 71; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

Dh 560 attattcattttatttttttcttagttataaaatag-taattctactaattaa 618  
||||||| || | ||||| || | ||||| ||| ||| |||||  
Qy 773 attatttcattttattttatttttggtgcataataacgaagtgtattattaa 832

Dh 619 aaaaaaaaaaaaaaaaaaaaaaaaaggaaaaa 651  
||||||| ||||| ||||| ||||| |||||  
Qy 833 aaaaaaaaaaaaaaaaaaaaaaaaagaaaaa 865

RESULT 6  
ID Q99393 standard; cDNA; 340 BP.  
AC Q99393;  
DT 08-FEB-1996 (first entry)  
DE Rat allograft inflammatory factor 2 (AIF-2) cDNA fragment.  
KW AIF-2; allograft inflammatory factor 2; transplant rejection;  
KX inhibitor; immunogenic; detection; diagnosis; ds.  
OS Rattus sp.  
PN WO9517506-A1.  
PR 29-JUN-1995.  
PF PD 21-DEC-1994; U14724.  
PR 21-DEC-1993; US-171385.  
PA (HARD ) HARVARD COLLEGE.  
PI Russell ME, Utans U;  
WP1: 95-240668/31.  
DR PT DNA encoding allograft rejection factors and immunogenic fragments -  
PT useful for identifying transplant rejection inhibitors  
CS Claim 11; Page 97; 138pp; English.  
PS Q99372-Q99393 are rat allograft inflammatory factor 2 (AIF-2) cDNA  
CC fragments. The AIF-2 gene is a differentially expressed allograft gene  
CC which is expressed in allograft tissue during transplant rejection.  
CC Identification of the rat AIF-2 product (R80321) or transcript  
CC indicates that allograft rejection is taking place. The rat AIF-2 gene,  
CC fragments of this gene and/or the gene product are therefore useful in  
CC the diagnosis of transplant rejection.  
CC The diagnostic methods used allow rejection (vascular inflammation)  
CC to be detected at an early stage and require only a small amount of  
CC biopsy material.  
CC (Sequence given in sequence ID listing in the specification gives  
CC nucleotide III as C)  
SQ Sequence 340 BP; 114 A; 38 C; 64 G; 124 T;

Query Match 5.08; Score 43; DB 16; Length 340;  
Best Local Similarity 78.7%; Pred. No. 1.67e-05;  
Matches 59; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Dh 4 ttttttttttttttttttttttttttttttttttacagtcatacacacagttatttt 63  
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Cp 865 tttttttttttttttttttttttttttttttttttaataaacctttattttatt 806

Dh 64 tagccataatgaaat 78  
| ||| || ||||  
Cp 805 tgaccacaataaat 791

RESULT 7  
ID Q99380 standard; cDNA; 350 BP.  
AC Q99380;

```

DR WPI: 92-217825/27.
DR P-PSDB: R25048.
PT Pathogen resistant plants prodn. using protein synthesis inhibiting
PT gene - or its fusion products, inserted into the genome, also
PT pharmaceutical use of derived synthesis inhibiting protein
PS Disclosure; Fig 3; 23pp; German.
CC The sequence is that of a gene encoding a protein which effectively
CC blocks protein synthesis by plant pathogens, e.g. Trichoderma reesi
CC and Fusarium sporotrichoides. It can be fused to active promoters
CC such as wun-1 and the fusion introduced into plant genetic material
CC to impart pathogen resistance to both mono- and di-cotyledonous
CC plants. The gene was isolated from mature barley seeds.
SQ Sequence 1078 BP; 276 A; 310 C; 321 G; 171 T;

Query Match 5.0%; Score 43; DB 4; Length 1078;
Best Local Similarity 85.2%; Pred. NO. 1.67e-05;
Matches 52; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1005 aaataaataataaataatgcagtttaaaaaaiaaaaaaaaaaaaaaaaaaaaaa 1064
Qy 805 aaataaataaaggttatttatttaaaaaaiaaaaaaaaaaaaaaaaaaaaaa 864

Db 1065 a 1065
Qy 865 a 865

RESULT 9
ID N70435 standard; cDNA; 1136 BP.
AC N70435;
DT 05-APR-1991 (first entry)
DE Sequence encoding insulin-like growth factor 1B (IGF-1B).
KW Growth promoter; lactation enhancer; cell proliferation; ss.
OS Homo sapiens.
PN EP-229750-A.
PD 22-JUL-1987.
PF 06-JAN-1987; 870001.
PR 07-JAN-1986; US-816662.
PR 20-NOV-1986; US-929671.
PA (NIW ) UNIV OF WASHINGTON.
PI Kivi GG, Rotwein PS;.
DR WPI: 87-200203/29.
PT New pre-pro-insulin-like growth factor-1 protein - obtd. by
PT recombinant DNA procedures for use as growth promoters for
PT enhancing lactation, for stimulating cell proliferation etc.
PS Example; Fig 5; 59pp; English.
CC A 42 base oligonucleotide corresponding to the DNA sequence encoding
CC amino acids 10 to 23 of mature human IGF-1 was synthesized (N70437).
CC The radiolabeled 42 mer was then employed to screen for IGF-1
CC containing DNA sequences in a human liver cDNA library. Insulin-
CC like growth factors-1A and -1B cDNAs were isolated from a human
CC library by using lambda gt 11 (N70435, N70436). The human IGF-1
CC genomic gene was isolated and mapped. It encodes at least two
CC preproinsulin-like growth factor-1 proteins. An essentially pure
CC proinsulin-like growth factor-1 protein comprising the sequence
CC of amino acids shown in Figure six is claimed (P70277).
SQ Sequence 1136 BP; 412 A; 230 C; 268 G; 226 T;

Query Match 5.0%; Score 43; DB 2; Length 1136;
Best Local Similarity 87.7%; Pred. NO. 1.67e-05;
Matches 50; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1073 aaataaataaagaatttaataaaaaaiaaaaaaaaaaaaaaaaaaaaaa 1129
Qy 809 aaataaataaaggttatttatttaaaaaaiaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 10
ID V15228 standard; cDNA; 1364 BP.
AC V15228;
DT 27-MAY-1998 (first entry)
DE Heterodimeric human interleukin 12 subunit 2 encoding cDNA.
KW Heterodimeric; human; interleukin 12; IL-12; vaccine; antigen;

```

---

```

KW pathogen; immune response; microorganism; cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 101..862
FT /tag= a
FT /product= "heterodimeric human interleukin 12 subunit"
PT US5723127-A.
PD 03-MAR-1998.
PF 25-MAR-1996; 621493.
PR 17-JUN-1994; US-265087.
PR 18-APR-1994; US-229282.
PR 25-MAR-1996; US-621493.
PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Scott P, Trinchieri G;
DR WPI: 98-205996/18.
DR P-PSDB; W44325.
PT Enhancing immune response to vaccine - comprises co-administering
PT interleukin-12 with antigenic composition.
PS Disclosure; Column 23-26; 23pp; English.
CC The present sequence encodes a heterodimeric human interleukin 12
CC (IL-12) subunit (putatively the 30 kD subunit) shown in the present
CC specification. The present invention describes a method for enhancing
CC the immune response elicited by an antigenic composition. The method
CC comprises administering IL-12 simultaneously or sequentially with the
CC antigenic composition. The IL-12 can be used in therapeutic
CC bacteria or viruses, especially in an amount of 0.1 mu g to 0.5 mg per
CC 25 mu g of antigen, or in cancer vaccines.
SQ Sequence 1364 BP; 446 A; 282 C; 275 G; 361 T;

Query Match 5.0%; Score 43; DB 40; Length 1364;
Best Local Similarity 90.6%; Pred. NO. 1.67e-05;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1299 aaataaagattgatttaaaaaaiaaaaaaaaaaaaaaaaaaaaaa 1351
Qy 813 aaataaggttatttatttaaaaaaiaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 11
ID T48099 standard; cDNA; 1364 BP.
AC T48099;
DT 21-MAR-1997 (first entry)
DE Human interleukin-12 30 kD subunit gene.
KW Human; interleukin 12; IL-12; heterodimeric complex; immunogenic; ds;
KW Leishmania; antigen; leishmaniasis; cell-mediated immunity; adjuvant;
KW Th1 helper cell; Th2 helper cell; alum; cytokine; bacterium; cancer;
KW vaccine; cell surface; membrane; glycoprotein; antigen presenting cell.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 101..862
FT /tag= a
FT /product= interleukin-12 30 kD subunit
PT US5571515-A.
PD 05-NOV-1996.
PF 18-APR-1994; 229282.
PR 18-APR-1994; US-229282.
PR 17-JUN-1994; US-265087.
PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Scott P, Trinchieri G;
DR WPI: 96-505347/50.
DR P-PSDB; W07399.
PT Immunogenic compsn. to improve cell mediated immune response -
PT contains soluble leishmania antigen and interleukin-12 as adjuvant
PS Disclosure; Column 23-26; 22pp; English.
CC This is the nucleotide sequence encoding the human interleukin (IL)-12
CC 30 kD subunit. IL-12 is a heterodimeric complex composed of the 30 kD
CC subunit and a 40 kD subunit (W07398). The complex is used in a novel
CC immunogenic composition comprising a soluble leishmania antigen with
CC IL-12, for protection against leishmaniasis. The addition of IL-12
CC improves cell-mediated immunity by inducing Th1 helper cells (as opposed

```

CC to TH2 cells which are induced by alum adjuvant) and does not cause  
 CC uncontrolled release of other cytokines (in contrast to bacterial  
 CC adjuvants). IL-12 can also be used as a cancer vaccine by association  
 CC with the protein B7, a soluble, cell-surface (membrane)-bound  
 CC glycoprotein which is expressed in antigen presenting cells.  
 SQ Sequence 1364 BP; 446 A; 282 C; 275 G; 361 T;

Query Match 5.0%; Score 43; DB 26; Length 1364;  
 Best Local Similarity 90.6%; Pred. No. 1.67e-05;  
 Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1299 aaataaagtattgaatataaaataaaataaaataaaataaaataaaataaa 1351  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 813 aaataagttattttataaaataaaataaaataaaataaaataaaataaa 865

RESULT 12  
 ID Q24371 standard; DNA; 1396 BP.  
 AC Q24371;  
 DE 16-SEP-1992 (first entry)  
 DE Natural killer cell stimulatory factor 30kd subunit.  
 KW NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;  
 KW IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;  
 KW ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_feature 14..1364  
 FT /tag= a  
 FT /label= p35nksf14-1-1  
 FT /note= "long clone containing 30kd NKSF subunit"  
 FT misc\_feature 146..1333  
 FT /tag= b  
 FT /label= p35nksf9-1-1  
 FT /note= "short clone containg 30kd NKSF subunit"  
 FT cds 113..873  
 FT /tag= c  
 FT /product= pro NKFS 30kd subunit  
 ET WO920256-A.  
 PN 02-APR-1992.  
 PF 04-SEP-1991; U06332.  
 PR 18-SEP-1990; US-584941.  
 PA (GENE-) GENETICS INST INC.  
 PA (WIST-) WISTAR INST.  
 PI Clark S C, Hewick R, Kobayashi M, Perussia B, Trinchieri G;  
 PI Wong G G;  
 DR WPI: 92-132124/16.  
 DR P-PSDB; R23729.  
 DE New natural killer cell stimulatory factor - useful as an  
 FT immunostimulant for inducing gamma-interferon and GM-CSF prodn.  
 FT in human peripheral blood lymphocytes  
 PS Claim 12; Page 23; 79pp; English.  
 CC NKSF is capable of inducing the production of gamma interferon in  
 CC human peripheral blood lymphocytes. It has subunits of 40 (R22769)  
 CC and 30-35 kd (R23729). It is claimed that NKSF is useful in the  
 CC treatment of bacterial and viral infections e.g. AIDS.  
 CC Degenerate probes were designed from tryptic fragments of NKSF  
 CC and used to screen a cDNA library made from poly A RNA from PdBu  
 CC induced 8866 cells to identify a cDNA encoding the small subunit  
 CC of NKSF.  
 CC See also R22769, Q23586, R23729.  
 SQ Sequence 1396 BP; 451 A; 294 C; 285 G; 366 T;

Query Match 5.0%; Score 43; DB 4; Length 1396;  
 Best Local Similarity 90.6%; Pred. No. 1.67e-05;  
 Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1312 aaataaagtattgaatataaaataaaataaaataaaataaaataaaataaa 1364  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 813 aaataagttattttataaaataaaataaaataaaataaaataaaataaa 865

RESULT 13  
 ID Q10896 standard; DNA; 1558 BP.

AC Q10896;  
 DT 13-MAY-1991 (first entry)  
 DE Encodes Xenopus Bone Morphogenetic Factor BMP-2B.  
 KW BMF; osteoporosis; fracture; cartilage; ss.  
 OS Xenopus laevis.  
 FH Key Location/Qualifiers  
 FT cds 105..1309  
 FT /tag= a  
 FT /product= BMP-2B  
 PN EP-416578-A.  
 PD 13-MAR-1991.  
 PF 05-SEP-1990; 117079.  
 PR 06-SEP-1989; JP-229250.  
 PR 20-JUL-1990; JP-190774.  
 PA (TAKE ) TAKEDA CHEMICAL IND KK.  
 PA (SCIT-) SCITECH RESEARCH CO.  
 PI Murakami K, Ueno N, Kato Y;  
 DR WPI: 91-075112/11.  
 DR P-PSDB; R10996.  
 DE Xenopus laevis bone morphogenetic protein and DNA encoding it -  
 FT used in therapy of fracture or osteoporosis  
 PS Claim 5; Fig 2; 28pp; English.  
 CC A Xenopus laevis unfertilised egg cDNA library in lambda gt10 was  
 CC screened with a PstI-HindIII fragment of X.laevis Xar14 chromosomal  
 CC DNA. Three clones were isolated, including clone Xbr23 which was  
 CC found to encode a protein having homology with X.laevis BMP-2B. They  
 CC were subcloned in pUC19 and used to transform competent E.coli HB101  
 CC cells. Transformant E.coli HB101/pXbr23 coding for the BMP-2B was  
 CC sequenced.  
 CC See also Q10890-5 and Q10897.  
 SQ Sequence 1558 BP; 504 A; 327 C; 348 G; 379 T;

Query Match 5.0%; Score 43; DB 2; Length 1558;  
 Best Local Similarity 85.2%; Pred. No. 1.67e-05;  
 Matches 52; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1467 aaaaaataaataagtcattatttttaaacataaaataaaataaaataaaataaa 1526  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 805 aaataaataaataaggtatttttttaaaataaaataaaataaaataaaataaa 864

Db 1527 a 1527  
 QY 865 a 865

RESULT 14  
 ID Q44391 standard; cDNA to mRNA; 3581 BP.  
 AC Q44391;  
 DT 14-SEP-1994 (first entry)  
 DE Sequence of murine OSF-4 cDNA.  
 KW OSF-4; cadherin; growth factor; osteogenesis; osteoblast; therapy;  
 KW diagnosis; ss.  
 OS Mus musculus.  
 FH Key Location/Qualifiers  
 FT cds 284..2671  
 FT /tag= a  
 PN EP-585801-A.  
 PD 09-MAR-1994.  
 PF 25-AUG-1993; 113602.  
 PR 28-AUG-1992; JP-230028.  
 PA (FARH ) HOECHST JAPAN LTD.  
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;  
 DR WPI: 94-076152/10.  
 DR P-PSDB; R49730.  
 DE New bone related, cadherin-like OSF-4 proteins - for treatment  
 FT and diagnosis of bone metabolic disease, and nucleic acid  
 FT encoding them  
 PS Claim 3; Page 13-17; 34pp; English.  
 CC cDNA libraries were constructed from the mouse osteoblastic cell  
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and  
 CC then as much common DNA as possible removed by hybridisation between  
 CC the 2 libraries. Residual EI-specific DNA was amplified, inserted  
 CC into lambda gt10 and screened by plaque hybridisation. A minibank of

Search completed: Tue Dec 15 19:01:29 1998  
Job time : 216 secs.

Query Match 5.0%; Score 43; DB 10; Length 3581;

RESULT 15  
ID V00423 standard; cDNA; 259 BP.  
AC V00423;  
AC V00423;  
DT 12-MAY-1998 (first entry)  
DE 3', fragment of clone H1075\_1.  
DE DT  
KW Human; secreted protein; molecular weight marker; genetic fingerprinting;  
KW antibody production; nutritional supplement; therapy; clone H1075\_1;  
KW peripheral blood mononucleocyte; PBMC; ds.  
OS Homo sapiens.  
PN W09740069-A2.  
PN W09740069-A2.  
PD 30-OCT-1997.  
PF 14-APR-1997; U06134.  
PF 19-APR-1996; US-635311.  
PR (GEM ) GENETICS INST INC.  
PI Jacobs K, Lavallie ER, Mccoy JW, Merberg D, Racie LA,  
PI Spaulding V;  
PI WPI: 97-535776/49.  
DR Isolated nucleic acid clones from ATCC 98028 encode novel secreted  
PT proteins - having many potential uses, e.g. as immunomodulators,  
PT cell proliferation or differentiation inhibitors or haematopoiesis  
PT regulators  
PS Claim 31: Page 74: 114pp: English

CC This sequence is the 3' portion of clone H1075\_1, which is a  
CC polynucleotide of the invention. This sequence was isolated from a human  
CC peripheral blood mononucleocyte (PBMC) cDNA library. The polynucleotide,  
CC which encodes a secreted protein, can be used, e.g. as a tissue or  
CC molecular weight marker, in genetic fingerprinting, to raise anti-protein  
CC or anti-DNA antibodies and in interaction trap assays. The protein can be  
CC used to assay biological activity, raise antibodies for use in  
CC immunoassays, as a marker, to identify inhibitors of its interactions and  
CC as a nutritional supplement. It may also have a very wide range of  
CC therapeutic and biological activities (no examples are given to support  
CC this), e.g. cytokine or modulator of cell proliferation and  
CC differentiation, immunostimulant or immunosuppressant, haematopoiesis  
CC regulator, bone, cartilage, tendon, ligament and/or nerve tissue growth  
CC stimulator, follicle inhibitor/stimulator, chemotactic/chemokinetic,  
CC homeostatic, thrombolytic or anti-inflammatory agent, antimicrobial,  
CC biorhythm, metabolism or behaviour modifier, anti-depressant or analgesic  
CC or psoriasis treatative.  
CC Sequence 259 BP: 105 A: 42 G: 75 T:  
SQ

Query Match 4.9%; Score 42; DB 38; Length 259;

Db 159 attnanttttgnataaaataa-aaatngttgttttgcccaaaaaaaaaaaaaaaa 217  
||| || | ||| ||| | ||| ||| ||| ||| |||  
Qv 791 attatttgtggtcaataataataaataagttatttttaaaaaaaaaaaaaaaaaa 850

```

Db      218 aaaaaaaaaaaaaa 232
        |||||
QY      851 aaaaaaaaaaaaaa 865

```

\*\*\*\*\*

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 19:01:47 1998; MasPar time 34.86 Seconds  
Tabular output not generated. 1157.525 Million cell updates/sec

Title: >US-09-109-864-1  
Description: (1-865) from US09109864.seq  
Perfect Score: 865  
N.A. Sequence: 1 ttgtggaggcagcagaga.....aaaaaaaaaaaaaaaa 865  
Comp: aaacacccctgcgtctct.....tttttttttttttttt

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 88822 seqs, 23323279 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-issued  
1:5\_COMB 2: PCT9\_COMB 3:backfiles1

Statistics: Mean 8.740; Variance 6.833; scale 1.279

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	44	5.1	7218	1	US-08-232-Sequence 14, Applicati	9.26e-09
C 2	43	5.0	340	1	US-08-171-Sequence 27, Applicati	2.65e-08
C 3	43	5.0	350	1	US-08-171-Sequence 14, Applicati	2.65e-08
C 4	43	5.0	1364	1	US-08-265-Sequence 3, Applicati	2.65e-08
C 5	43	5.0	1558	1	US-08-455-Sequence 7, Applicati	2.65e-08
C 6	43	5.0	7218	1	US-08-232-Sequence 14, Applicati	2.65e-08
C 7	42	4.9	3073	2	PCT-US91-0Sequence 30, Applicati	7.52e-08
C 8	42	4.9	3073	2	PCT-US91-0Sequence 31, Applicati	7.52e-08
C 9	41	4.7	198	2	PCT-US92-1Sequence 16, Applicati	2.13e-07
C 10	41	4.7	688	3	5498694-3Patent No. 5498694	2.13e-07
C 11	41	4.7	1046	1	US-08-361-Sequence 4, Applicati	2.13e-07
C 12	41	4.7	2389	2	PCT-US96-1Sequence 1, Applicati	2.13e-07
C 13	41	4.7	3333	1	US-08-199-Sequence 1, Applicati	2.13e-07
C 14	41	4.7	3933	2	PCT-US95-0Sequence 1, Applicati	2.13e-07
C 15	41	4.7	5108	1	US-07-642-Sequence 2, Applicati	5.98e-07
C 16	40	4.6	347	1	US-08-104-Sequence 1, Applicati	5.98e-07
C 17	40	4.6	1283	1	US-08-174-Sequence 19, Applicati	5.98e-07
C 18	40	4.6	2158	1	US-08-261-Sequence 1, Applicati	5.98e-07
C 19	40	4.6	2158	1	US-07-602-Sequence 1, Applicati	5.98e-07
C 20	40	4.6	2550	3	5258287-23Patent No. 5258287	5.98e-07

21	40	4.6	2940	3	5171671-1	Patent No. 5171671	5.98e-07
22	40	4.6	3214	1	US-08-484-Sequence 17, Applicati	Sequence 17, Applicati	5.98e-07
23	40	4.6	3214	1	US-08-484-Sequence 5, Applicati	Sequence 5, Applicati	5.98e-07
24	39	4.5	140	1	US-08-628-Sequence 6, Applicatio	Sequence 6, Applicatio	1.67e-06
25	39	4.5	240	1	US-08-628-Sequence 3, Applicati	Sequence 3, Applicati	1.67e-06
26	39	4.5	289	1	US-08-341-Sequence 11, Applicati	Sequence 11, Applicati	1.67e-06
27	39	4.5	654	2	PCT-US95-0Sequence 24, Applicati	Sequence 24, Applicati	1.67e-06
28	39	4.5	1493	1	US-08-593-Sequence 8, Applicatio	Sequence 8, Applicatio	1.67e-06
29	39	4.5	1493	1	US-08-340-Sequence 24, Applicati	Sequence 24, Applicati	1.67e-06
30	39	4.5	1641	1	US-08-300-Sequence 1, Applicatio	Sequence 1, Applicatio	1.67e-06
31	39	4.5	1747	1	US-08-176-Sequence 1, Applicatio	Sequence 1, Applicatio	1.67e-06
32	39	4.5	1875	2	PCT-US96-1Sequence 1, Applicatio	Sequence 1, Applicatio	1.67e-06
33	39	4.5	1897	1	US-08-184-Sequence 1, Applicatio	Sequence 1, Applicatio	1.67e-06
34	39	4.5	2010	1	US-07-864-Sequence 4, Applicatio	Sequence 4, Applicatio	1.67e-06
35	39	4.5	2628	1	US-08-143-Sequence 1, Applicatio	Sequence 1, Applicatio	1.67e-06
36	39	4.5	2760	1	US-08-101-Sequence 3, Applicatio	Sequence 3, Applicatio	1.67e-06
C 37	39	4.5	3238	2	PCT-US94-1Sequence 5, Applicatio	Sequence 5, Applicatio	1.67e-06
C 38	39	4.5	5852	1	US-07-867-Sequence 2, Applicatio	Sequence 2, Applicatio	1.67e-06
39	38	4.4	277	1	US-08-244-Sequence 18, Applicati	Sequence 18, Applicati	4.63e-06
40	38	4.4	399	1	US-08-298-Sequence 13, Applicati	Sequence 13, Applicati	4.63e-06
41	38	4.4	635	2	PCT-US94-0Sequence 35, Applicati	Sequence 35, Applicati	4.63e-06
42	38	4.4	1393	1	US-08-174-Sequence 18, Applicati	Sequence 18, Applicati	4.63e-06
43	38	4.4	4104	1	US-07-998-Sequence 94, Applicati	Sequence 94, Applicati	4.63e-06
44	38	4.4	6671	2	PCT-US95-0Sequence 1, Applicatio	Sequence 1, Applicatio	4.63e-06
45	38	4.4	6671	1	US-08-457-Sequence 1, Applicatio	Sequence 1, Applicatio	4.63e-06

ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DE  
DT  
CC Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22313-0299  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313  
CC FILING DATE:  
CC APPLICATION NUMBER: EP 91 114 300.6  
CC FILING DATE: 26-AUG-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.  
CC REGISTRATION NUMBER: 29,768  
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300  
CC TELEFAX: (703)683-4109  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:

```
CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: PT9gpt-F1s
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match      5.1%; Score 44; DB 1; Length 7218;
Best Local Similarity 0.7%; Pred. No. 9,26e-09;
Matches 2; Conservative 165; Mismatches 123; Indels 0; Gaps 0;

Db 1147 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1206
Cp 865 tttttttttttttttttttttttttttttttttttttaataaaccttatttattt 806
Db 1207 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1266
Cp 805 tgaccacaaataataagaataaatgccacattcttaagccattgtattt 746
Db 1267 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1326
Cp 745 tggagtacttgttagctcaaacacctgcctcgtgagtcagtcaggccctgaagg 686
Db 1327 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1386
Cp 685 cagatggagccatccccaggtcctcgttgaggaggTgcagagtcgaggaaccttgg 626
Db 1387 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1436
Cp 625 cactctgtgctccctcaggccccaagctcctgtagactcagctcgt 576

RESULT 2
ID US-08-171-385-27 STANDARD; DNA; UNC; 340 BP.
AC xxxxxx
DT
DE Sequence 27, Application US/08171385
CC Sequence 27, Application US/08171385
CC Patent No. 5527884
CC GENERAL INFORMATION:
CC APPLICANT: Mary E. Russell
CC APPLICANT: Ulrike Utans
CC TITLE OF INVENTION: Mediators of Chronic Allograft
CC TITLE OF INVENTION: Rejection
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 MB
CC COMPUTER: IBM PS/2 Model 50z or 55sx
CC OPERATING SYSTEM: MS-DOS (Version 5.0)
CC SOFTWARE: Wordperfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/171.385
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 05433/006001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
```

RESULT 6  
ID US-0-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
DT  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.

Sequence 7, Application US/08455550  
Sequence 7, Application US/08455550  
Patent No. 5670338  
GENERAL INFORMATION:  
APPLICANT: MURAKAMI, KAZUO  
APPLICANT: UENO, NAOTO  
APPLICANT: KATO, YUKIO  
TITLE OF INVENTION: XENOPUS LAEVIS BONE MORPHOGENETIC PROTEINS AND USE THE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dike, Bronstein, Roberts & Cushman  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,550  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/056,564  
FILING DATE: 30-APR-1993  
APPLICATION NUMBER: 07/577,892  
FILING DATE: 05-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Eisenstein, Ronald I  
REGISTRATION NUMBER: 30628  
REFERENCE/DOCKET NUMBER: 40302-FWC-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX: 200291  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1558 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
SEQUENCE 1558 BP; 504 A; 328 C; 347 G; 379 T; 0 OTHER.

Query Match 5.0%; Score 43; DB 1; Length 1558;  
Best Local Similarity 85.2%; Pred. No. 2.65e-08;  
Matches 52; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Ddb 1467 AAAAAAAAAAATAAGTCATTTATTTTAAACATAAAAAAAAAAAAAAAAAAAAAAAA 1526  
||| ||||| | ||| ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
Qy 805 aaataacatacaagggtatttattttcaaaaaaaaaaaaaaaaaaaaaaa 864

Ddb 1527 A 1527  
|  
Qy 865 a 865

RESULT 6  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx  
Dt  
DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367  
CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.

[illegible]



QY 865 a 865

RESULT 11  
ID US-08-361-467B-4 STANDARD; DNA; UNC; 1046 BP.  
AC xxxxxx  
DT  
DE Sequence 4, Application US/08361467B  
CC Sequence 4, Application US/08361467B  
CC Patent No. 5633441  
CC GENERAL INFORMATION:  
CC APPLICANT: De Greef, Willy  
CC APPLICANT: Van Emmelo, John  
CC APPLICANT: De Oliveira, Dulce E.  
CC APPLICANT: De Souza, Maria-Helena  
CC APPLICANT: Van Montagu, Marc  
CC TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR  
CC NUMBER OF INVENTION: EMBRYOS  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
CC STREET: P.O. Box 1404  
CC CITY: Alexandria  
CC STATE: Virginia  
CC COUNTRY: United States  
CC ZIP: 22313-1404  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/361,467B  
CC FILING DATE: 22-DEC-1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/681,492  
CC FILING DATE: 04-APR-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/EP90/01275  
CC FILING DATE: 01-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: EP 89 402 224.3  
CC FILING DATE: 04-AUG-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Schulman, Robert M.  
CC REGISTRATION NUMBER: 31,196  
CC REFERENCE/DOCKET NUMBER: 010830-027  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703) 836-6620  
CC TELEFAX: (703) 836-2021  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1046 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: double  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cdna to mRNA  
CC IMMEDIATE SOURCE:  
CC CLONE: 3C9  
CC SEQUENCE 1046 BP; 360 A; 254 C; 166 G; 266 T; 0 OTHER.

Query Match 4.7%; Score 41; DB 1; Length 1046;  
Best Local Similarity 75.9%; Pred. No. 2.13e-07;  
Matches 60; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Db 948 ATTATAGTTTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1007  
QY 787 attatttttggtgcataataataaggttatttatttataaaaaa 846  
Db 1008 AAAAAAAAAAAAAAAAAA 1026  
QY 847 aaaaaaaaaaaaaa 865

RESULT 12  
ID PCT-US96-12860-1 STANDARD; DNA; UNC; 2589 BP.  
AC xxxxxx  
DT  
DE Sequence 1, Application PC/TUS9612860  
CC Sequence 1, Application PC/TUS9612860  
CC GENERAL INFORMATION:  
CC APPLICANT: TULARIK, INC.  
CC TITLE OF INVENTION: INHIBITORS OF APOPTOSIS  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
CC STREET: 4 Embarcadero Center, Suite 3400  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US96/12860  
CC FILING DATE: 06 AUG 1996  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Brezner, David J.  
CC REGISTRATION NUMBER: 24,774  
CC REFERENCE/DOCKET NUMBER: A-62464/DJB  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415)781-1989  
CC TELEFAX: (415)398-3249  
CC INFORMATION FOR SEQ ID NO: 1:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2589 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cdna  
CC SEQUENCE 2589 BP; 844 A; 436 C; 522 G; 787 T; 0 OTHER.  
Query Match 4.7%; Score 41; DB 2; Length 2589;  
Best Local Similarity 83.6%; Pred. No. 2.13e-07;  
Matches 51; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
Db 2510 AACATCTCATATAAGTCCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2569  
QY 805 aaataataataaggttatttatttataaaaaa 864  
Db 2570 A 2570  
QY 865 a 865  
RESULT 13  
ID US-08-199-776-1 STANDARD; DNA; UNC; 3933 BP.  
AC xxxxxx  
DT  
DE Sequence 1, Application US/08199776  
CC Sequence 1, Application US/08199776  
CC Patent No. 5594120  
CC GENERAL INFORMATION:  
CC APPLICANT: Brenner, Michael B.  
CC APPLICANT: Parker, Christina M.  
CC TITLE OF INVENTION: NO. 5594120el integrin alpha subunit  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:

COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02044  
FILING DATE: herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/199,776  
FILING DATE: 18 February 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3933 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: mucosal lymphocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 126..3662  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 180..3659  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 126..179  
SEQUENCE 3933 BP: 938 A; 1062 C; 1074 G; 859 T; 0 O

Query Match 4.7%; Score 41; DB 2; Length 3933;  
Best Local Similarity 78.9%; Pred. No. 2.13e-07;  
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 3847 ATTTGGPATTAAATAAAATTACACTTTCTTTGCAAAAAAATAAAAAAAAAAAAAA 3906

Qy 795 atttgggtcaataaataaataaaggttatttatttaaataaaaaaataaaaaaataaaa 854

```

Db 3907 AAAAAAAAAA 3917
    |||
QY 855 aaaaaaaaaa 865

```

RESULT 15  
ID US-07-642-002-1 STANDARD; DNA; UNC; 5108 BP.  
AC xxxxxx

DE Sequence 1, Application US/07642002  
 CC Sequence 1, Application US/07642002  
 CC Patent No. 5268465  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Brett, David S.  
 CC APPLICANT: Hwang, Paul M.  
 CC APPLICANT: Reed, Randall  
 CC APPLICANT: Snyder, Solomon H.  
 CC TITLE OF INVENTION: Purification and Molecular Cloning of Nitric  
 CC  
 CC TITLE OF INVENTION: Oxide Synthase  
 CC  
 CC NUMBER OF SEQUENCES: 2  
 CC  
 CC CORRESPONDENCE ADDRESS:  
 CC



\*\*\*\*\*

W P E R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 15:53:21 1998; MasPar time 5.62 Seconds  
Tabular output not generated. 382.814 Million cell updates/sec

Title: >US-09-109-864-2

Description: (1-133) from US09109864.pep

Perfect Score: 980

Sequence: 1 MSPRLEVPCHALPQGLSPG.....LEQLRELIRISGVOLVCVHS 133

Scoring table: PAM 150  
Gap 11

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq32

8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 30.678; Variance 131.480; scale 0.233

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	140	14.3	317	19 W03519	Prostate carcinoma tu	2.28e-03
2	109	11.1	316	13 R75702	Rat galectin-8.	7.19e-01
3	92	9.4	278	13 R75997	Ostertagia protective	1.41e-01
4	89	9.1	323	29 W46883	Protein sequence enco	2.34e-01
5	89	9.1	323	23 W11841	Human galectin-4-like	2.34e-01
6	89	9.1	323	20 W05551	Human colon specific	2.34e-01
7	89	9.1	343	26 W36070	E. coli DNA polymeras	2.34e-01
8	89	9.1	983	13 R75701	BAV3 ORF6-encoded pro	2.34e-01
9	87	8.9	343	8 R40125	DNA polymerase III ho	3.27e-01
10	85	8.7	311	28 W35129	Snpr activator protei	4.56e-01
11	85	8.7	495	2 R07143	Neuronal nicotinic ac	4.56e-01
12	84	8.6	212	10 R54922	Cdil polypeptide.	5.38e-01
13	84	8.6	704	19 W01102	Signal transducer and	5.38e-01
14	84	8.6	748	19 W01101	Signal transducer and	5.38e-01
15	83	8.5	264	3 R12531	Mac2.16 expression pr	6.34e-01
16	83	8.5	278	3 R13338	Mac-2 protein includi	6.34e-01
17	83	8.5	687	23 W22458	Yeast wild-type aceto	6.34e-01
18	83	8.5	687	1 P81152	Herbicide sensitive,	6.34e-01

19 83 8.5 863 13 R78233 Chicken oocyte recept 6.34e-01  
20 81 8.3 123 8 R43827 Anti-lysozyme VH. 8.79e-01  
21 81 8.3 847 15 R88320 IL-4 Stat peptide. 8.79e-01  
22 81 8.3 1420 21 W13826 Yeast transcription r 8.79e-01  
23 81 8.3 15281 9 R44929 T. niveum Cyclosporin 8.79e-01  
24 80 8.2 379 11 R56870 Canine gastric lipase 1.03e-02  
25 80 8.2 379 23 W09382 Dog gastric lipase pr 1.03e-02  
26 80 8.2 380 11 R56871 Canine gastric lipase 1.03e-02  
27 80 8.2 785 22 W16303 Yeast mitochondrial CO 1.03e-02  
28 79 8.1 227 19 W05530 Mouse TRADD intracell 1.21e-02  
29 79 8.1 357 11 R60255 Human NF-ATP. 1.21e-02  
30 79 8.1 570 5 R09316 Pyruvate oxidase (p17 1.21e-02  
31 79 8.1 570 5 R05793 Pyruvate oxidase (w1 1.21e-02  
32 79 8.1 921 19 W02248 Human transcription f 1.21e-02  
33 79 8.1 1484 2 R11749 Human alpha-2 macrogl 1.21e-02  
34 78 8.0 119 15 R82986 LK26 humanised VH reg 1.42e-02  
35 78 8.0 125 8 R42831 Cytochrome P450 homol 1.42e-02  
36 78 8.0 135 1 P91964 HL-60 and placenta le 1.42e-02  
37 78 8.0 135 2 R08153 Human lectin gene pro 1.42e-02  
38 78 8.0 135 10 R52745 Amino acid sequence 1.42e-02  
39 78 8.0 135 2 R22625 Human GBP. 1.42e-02  
40 78 8.0 184 22 W20635 H. pylori transmembra 1.42e-02  
41 78 8.0 195 9 R47339 Peptide fragment of m 1.42e-02  
42 78 8.0 677 23 W25017 TATA-binding protein 1.42e-02  
43 78 8.0 677 11 R56492 TATA-binding protein- 1.42e-02  
44 78 8.0 677 20 W06083 Human TATA-binding pr 1.42e-02  
45 78 8.0 686 5 R25591 RING1 antigenic pept 1.42e-02

#### ALIGNMENTS

RESULT 1  
ID W03519 standard; Protein; 317 AA.

AC W03519;

DE 26-NOV-1996 (first entry)

DT Prostate carcinoma tumour antigen.

KW Prostate carcinoma tumour antigen; PCTA-1; oncogene; cancer

KW diagnosis; therapy; metastasis; cell surface expression;

KW surface epitope masking; galectin-8.

OS Homo sapiens.

PN W09621671-A1.

PD 18-JUL-1996.

PF 11-JAN-1996; U00307.

PR 11-JAN-1995; US-371377.

PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Shen R;

DR WPI: 96-342235/34.

DR N-PSDB; T37414.

PT DNA encoding prostate carcinoma tumour antigen and prostate tumour

PT inducing genes - and related proteins, vectors, antibodies, etc.,

PT for diagnosis and treatment of metastatic cancer

PS Disclosure: Fig 18A; 169pp; English.

CC PCTA (W03519) that is a putative tumour-associated antigen present

CC on prostate cancer cells. Its amino acid sequence was deduced

CC from a cDNA clone (T37414) isolated from human prostatic carcinoma

CC LNCap cells using a surface epitope masking (SEM) technique. SEM

CC is a method for producing MAb specific for cell surface-expressed

CC molecules. PCTA is a new member of the galectin family,

CC galectin-8. Antibodies raised against PCTA are useful for the

CC detection of metastatic cancer. Cpds. that inhibit PCTA can be

CC used in treatment of cancer.

SQ Sequence 317 AA;

Query Match 14.3%; Score 140; DB 19; Length 317;

Best Local Similarity 24.6%; Pred.No. 2.28e-03;

Matches 34; Conservative 41; Mismatches 56; Indels 7; Gaps 5;

Db 180 tpglsipfaarintpmpgrrvvgvevnanaksnfvdlagkskdialhlnprlnkaf 239.

QY 2 SPRLVPCSHALPQGLSPGVIIIRGLVQLQPKHFTVSL-ROQAAHAPVTLRASFDRTL 60

Db 240 vnsflqesgweernitsfpfpngmyfemilycdvrefkvavngvhsleykhrfke-ls 238

[illegible]

```

AC W1841;
DE 20-OCT-1997 * (first entry)
DE Human galectin-4-like protein.
KW galactin-4; lactose binding; stomach; intestine; reagent; drug;
KW research; carbohydrate.
OS Homo sapiens.
PN WO9703190-AI.
PD 30-JAN-1997.
PF 09-JUL-1996; J01899.
PR 11-JUL-1995; JP-174778.
PP (SAGA ) SAGAMI CHEM RES CENTRE.
PI Kamata K, Kato S, Sekine S, Yamaguchi T;
DR WPI: 97-119046/11.
DR N-PSTDB; T59539.
DR Recombinant human galectin 4-like protein and gene - has
PT lactose-binding ability, used in research involving carbohydrate(s)
PS Claim 1: Page 13-16; 25pp; Japanese.
CC This sequence is a human galectin 4 (a lactose-binding protein)-like
CC protein. The protein, which is expressed specifically in the human
CC stomach and intestines, binds lactose and is useful as a drug and as a
CC reagent for research involving carbohydrates.
SQ Sequence 323 AA;

Query Match          9.1%; Score 89; DB 23; Length 323;
Best Local Similarity 24.8%; Pred. No. 2.34e+01;
Matches 33; Conservative 36; Mismatches 55; Indels 9; Gaps 5;

Db      192 vpyfgrlqggltarrrtiikgyvptgtksfaifkvsgsdialhinprmgngvtvrnsI 251
QY      7 VPCSHALPGLSPGVQVIIVRGLVLQPKFTVSLR-DQAHAHVTVIRASFADRTILAW--- 62

Db      252 lngswgeekktthmpfg--pgoffdlsrcldrfkvyangqhlfdfahrIsafgrdvt 309
QY      :: || :||: || |::|:: ||: ||::|: ||: ||::|: ||: ||::|: |:::
        63 IS-RWG-QKKLISAPFLVPORFFEVLLLFQSGGLKALNGGCGLGATSNMQALEQLRE 119

Db      310 leiaggdvttlsyqq 322
QY      | | | | | | | | | |
        120 LRISSGVQLYCVH 132

RESULT 6
ID W06551 standard; Protein; 323 AA.
AC W06351;
DT 13-MAR-1997 (first entry)
DE Human colon specific gene CSG9 complete polypeptide.
KE Colon specific gene; CSG9; colon cancer; metastasis; diagnosis;
KW therapy; antibody; vaccine.
OS Homo Sapiens.
PN WO9639419-AI.
PD 12-DEC-1996.
PF 06-JUN-1995; U07289.
PR 06-JUN-1995; WO-U07289.
PP (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Yu G;
DR WPI: 97-043054/04.
DR N-PSTDB; T45888.
DR Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
PT Claim 8; Fig 9; 60pp; English.
CC Novel polypeptides (W06345-53) are encoded by cDNA clones (see also
CC T45980-92) corresponding to 13 human colon specific genes,
CC designated CSG1, CSG2, etc., that are primarily expressed in
CC tissues derived from the colon. Recombinant CSG polypeptides can
CC be produced in transformed host cells. They are useful diagnostic
CC markers for colon cancer and for colon cancer metastasis and can
CC also be used to screen for (ant)agonist cpds. of therapeutic or
CC diagnostic value. Antibodies raised against the colon-specific
CC polypeptides may be used to target colon cancer cells or as part
CC of a colon cancer vaccine.
SQ Sequence 323 AA;

Query Match          9.1%; Score 89; DB 20; Length 323;

```

FT modified\_site /label= N-glycosylation\_site  
 FT 257..259  
 FT /label= N-glycosylation\_site  
 FT 268..270  
 FT modified\_site /label= N-glycosylation\_site  
 FT 426..428  
 FT modified\_site /label= N-glycosylation\_site  
 FT 574..576  
 FT modified\_site /label= N-glycosylation\_site  
 FT 714..716  
 FT modified\_site /label= N-glycosylation\_site  
 FT 819..821  
 FT modified\_site /label= N-glycosylation\_site  
 FT 847..849  
 FT modified\_site /label= N-glycosylation\_site  
 FT 853..855  
 FT modified\_site /label= N-glycosylation\_site  
 FT 894..896  
 FT modified\_site /label= N-glycosylation\_site  
 FT 913..915  
 FT modified\_site /label= N-glycosylation\_site  
 FT 973..975  
 FT /label= N-glycosylation\_site  
 PN W09516048-A.  
 PD 15-JUN-1995.  
 PF 09-DEC-1994; CA0678.  
 PR 09-DEC-1993; US-164292.  
 PA (UYSA-) UNIV SASKATCHEWAN.  
 PI Babink LA., Graham FL, Mittal SK, Prevec L;  
 DR WPI: 95-224330/29.  
 DR N-PSDB: Q90769.  
 PT Bovine adenovirus expression vector system - comprising insertion of  
 PT required genes into deletions in E1 and E3 and transformation of  
 PT mammalian cell lines, useful in vaccines and gene therapy  
 PS Disclosure; Fig 7H-R; 159pp; English.  
 CC BAV3 strain WBR-1 genomic DNA, from between 77 and 92 m.u.,  
 CC was cloned into plasmids and sequenced. ORFs that had the  
 CC potential to encode polypeptides of at least 50 amino acids  
 CC (R75756-61) were identified. The amino acid sequence at the  
 CC N-terminus of the ORF6-encoded protein shared approx. 60%  
 CC identity with the Hsd2 fiber protein tail, but there was little  
 CC similarity in the knob region.  
 SQ Sequence 983 AA;  
 Query Match 9.1%; Score 89; DB 13; Length 983;  
 Best Local Similarity 32.6%; Pred. No. 2.34e+01;  
 Matches 14; Conservative 12; Mismatches 15; Indels 2; Gaps 2;  
 Db 341 ltfengavrakl-gpglgtddsgsvvrtgrglrvangqvqif 382  
 QY I::I::: I::I::: : : I::I::: I::I:::  
 88 LLFQEGGLKALNGQLGATSMNQALQLRELRLIS-GSVQLY 129  
 RESULT 9  
 ID R40125 standard; Protein; 343 AA.  
 AC R40125;  
 DE 26-JAN-1994 (first entry)  
 DT DNA polymerase III holoenzyme delta subunit hola protein.  
 KW Enzyme; Chromosomal replicase.  
 OS Escherichia coli.  
 PN W09315115-A.  
 PD 05-AUG-1993.  
 PF 22-JAN-1993; U00627.  
 PR 24-JAN-1992; US-826926.  
 PA (CORR ) CORNELL RES FOUND INC.  
 PI ODonnell ME;  
 DR WPI: 93-258618/32.  
 DR N-PSDB: Q47197.  
 PT Nucleotide sequences for DNA polymerase III holo enzyme sub-units  
 PT - enzyme with 5 or 6 sub-units having same activity as 10 subunit  
 PT enzyme  
 PS Example; Page 16-17; 115pp; English.  
 CC The sequence is that encoded by the DNA polymerase III holoenzyme

CC delta subunit hola gene.  
 SQ Sequence 343 AA;  
 Query Match 8.9%; Score 87; DB 8; Length 343;  
 Best Local Similarity 39.0%; Pred. No. 3.27e+01;  
 Matches 23; Conservative 13; Mismatches 16; Indels 7; Gaps 6;  
 Db 231 hllqqlrlegsepvlrrtl-qrel-lllwnlkrqsahtp--lraif-dkhrvwqnrrg 284  
 QY I::I::: I::I::: I::I::: I::I::: I::I::: I::I::: I::I::: I::I:::  
 11 HALPQ-GLSPGQ-VIIIVGLVQLQEPKHFTVSLRDQAHAHAPVTLRASFDRTLAWISRWG 67  
 RESULT 10  
 ID W36129 standard; Protein; 311 AA.  
 AC W36129;  
 DE 21-MAY-1998 (first entry)  
 DT SnPR activator protein.  
 KW Daunomycin C-14 hydroxylase; doxa gene; doxorubicin; daunomycin;  
 KW 13-dihydrocarminomycin; carminomycin; anthracycline;  
 KW anticancer; cytostatic; cancer; therapy; plasmid pANT195; SnPR.  
 OS Streptomyces sp.  
 FH Key Location/Qualifiers  
 FT Domain 18..39  
 FT /note= "alpha-helix-beta-turn-alpha helix  
 FT DNA-binding domain"  
 FT W09744439-A2.  
 PN 27-NOV-1997.  
 PD 22-MAY-1997; U08690.  
 PR 24-MAY-1996; US-853650.  
 PA (OHIS ) UNIV OHIO STATE RES FOUND.  
 PI DeSanti CL, Dickens ML, Strohl WA;  
 DR WPI: 98-018495/02.  
 DR N-PSDB: V01451.  
 PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase  
 PT - also hydroxylation and oxidation of other anthracycline(s) with  
 PT the same enzyme  
 PS Disclosure; Fig 6; 59pp; English.  
 CC This protein is the encoded product of the Streptomyces snpr  
 CC activator gene. The snpr activator gene is incorporated in  
 CC novel plasmid pANT195 (see V01451) that also includes the  
 CC snpr-activated snpa promoter and a modified doxa gene (see  
 CC V01447) of Streptomyces sp. strain C5. The doxa gene codes for  
 CC daunomycin C-14 hydroxylase (see W36128), an enzyme capable of  
 CC converting daunomycin to the anticancer agent doxorubicin. Host  
 CC cells, especially Streptomyces host cells, transformed with pANT195  
 CC can be used in methods for the production of doxorubicin from  
 CC daunomycin or for the hydroxylation and oxidation of other  
 CC anthracyclines.  
 SQ Sequence 311 AA;  
 Query Match 8.7%; Score 85; DB 28; Length 311;  
 Best Local Similarity 33.3%; Pred. No. 4.56e+01;  
 Matches 21; Conservative 14; Mismatches 22; Indels 6; Gaps 5;  
 Db 136 hldvafvhevgslrvpeglrvrvlqvrep-qf-vcl--padh-paaeatasyasptw-t 189  
 QY I::I::: I::I::: I::I::: I::I::: I::I::: I::I::: I::I::: I::I:::  
 4 RLEVPCHALPQGLSPGQVIVRGLVQLQEPKHFTVSLRDQAHAHAPVTLRASFDRTLAWI 63  
 Db 190 trw 192  
 QY I::I:::  
 64 SRW 66  
 RESULT 11  
 ID R07143 standard; protein; 495 AA.  
 AC R07143;  
 DT 24-JAN-1991 (first entry)  
 DE Neuronal nicotinic acetylcholine receptor beta 4 subunit.  
 KW Rat; nAChR.  
 OS Rattus rattus.  
 FH Key Location/Qualifiers  
 FT peptide 1..20  
 FT /label=signal peptide

[illegible]

```

RESULT 14
ID W01101 standard; Protein; 748 AA.
AC W01101;
DT 03-JAN-1997 (first entry)
DE Signal transducer and activator of transcription 4-alpha.
KW Signal transducer and activator of transcription 4; Stat4;
KW Interleukin-12; signal transduction; neoplasm; cancer; diagnosis;
KW therapy.
OS Homo sapiens.
PN W09629341-A1.
PD 26-SEP-1996.
PF 22-MAR-1996; U03868.
PR 22-MAR-1995; US-408318.
PA (TULA-) TULARIK INC.
PI Hoey T, Rothe M;
DR WPI; 96-443132/44.
DR N-PSDB; T38324.
PT DNA encoding human signal transducer and activator of transcription
PT 4 - used for screening for drugs useful in diagnosis, prognosis and
PT treatment of e.g. neoplasia
PS Claim 1; Page 21-23; 32pp; English.
CC Human signal transducer and activator of transcription 4-alpha
CC (Stat4-alpha) (W01101) is an interleukin-12 signal transducer
CC protein that binds an intracellular domain of an interleukin-12
CC receptor subunit. Its amino acid sequence was deduced from a cDNA
CC clone (T38324) isolated from a Jurkat cDNA library. A second,
CC truncated isoform, Stat4-beta (W01102), is the product of a cDNA
CC clone (T38325) obt'd. human peripheral blood lymphocytes.
CC Recombinant Stat4 proteins can be produced in transformed host
CC cells. They find use in screening assays for cpds. useful in the
CC diagnosis, prognosis or treatment of diseases associated with
CC undesirable cell growth, differentiation and/or cytokine signal
CC responsiveness.
SQ Sequence 748 AA.

Query Match 8.6%; Score 84; DB 19; Length 748;
Best Local Similarity 23.8%; Pred. No. 5.38e+01;
Matches 19; Conservative 27; Mismatches 30; Indels 4; Gaps 4;

Db 522 ltvqsysdghltw-akfckehlpkgsffwt-wleaidlkhkhlplwidymgfv 579
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 49 VTLRASFDRTLAWISRWGOKLISAPFLFPQRFVLLLFQEGGLKALNGQGLGATS 108
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 580 kekerl-llkd-kmpgtfll 597
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 109 MNQQALEQLRELRISGVQL 128

RESULT 15
ID R12531 standard; Protein; 264 AA.
AC R12531;
DT 12-SEP-1991 (first entry)
DE Mac2.16 expression product.
KW HMEBP; leishmaniasis; Mouse Mac-2; laminin.
OS Mus musculus.
PN W09108290-A.
PD 13-JUN-1991.
PF 29-NOV-1990; U06948.
PR 30-NOV-1989; US-444195.
PR 14-SEP-1990; US-582628.
PA (GEHO-) GEN HOSPITAL CORP.
PI Pillai S, Cherayil BJ;
DR WPI; 91-193196/26.
DR N-PSDB; Q12207.
PT Recombinant gene encoding human macrophage carbohydrate -
PT Ige-binding protein and antibody used to treat, diagnose and
PT prevent e.g. inflammatory bowel disorder, leishmaniasis, hayfever
PT and bronchial
PS Disclosure; Fig 1; 36pp; English.
CC Clone Mac2.16 was isolated from a lambda gt11-p388b1 expression
CC library prepared from activated mouse macrophages. This is the

```

```

CC deduced amino acid sequence of the Mac2.16 expression product. It
CC has predicted mol. wt. 27,482. The N-terminal domain contains a
CC repetitive Proline/Glycine-rich motif and the C-terminal domain
CC contains a sequence found in S-type lectins.
CC See also Q12208-Q12211.
SQ Sequence 264 AA;

Query Match 8.5%; Score 83; DB 3; Length 264;
Best Local Similarity 27.0%; Pred. No. 6.34e+01;
Matches 17; Conservative 15; Mismatches 30; Indels 1; Gaps 1;

Db 195 wg-keerqsafpfesgkpfkiqlvleadhfkavndahllqynhrmknleisqlgsd 253
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 66 WGOKKLISAPFLFPQRFVLLLFQEGGLKALNGQGLGATSMNQQALEQLRELRISGS 125
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 254 itl 256
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 126 VQL 128
: : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Search completed: Tue Dec 15 15:54:12 1998
Job time : 51 secs.

```

\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 15:56:27 1998; MasPar time 2.20 Seconds  
Tabular output not generated. 427.387 Million cell updates/sec

Title: >US-09-109-864-2  
Description: (1-133) from US09109864.pep  
Perfect Score: 980  
Sequence: 1 MSPRLVPCSHALPQGLSPG.....LEQLRELRLISGVOLXCVHS 133

Scoring table: PAM 150

Gap 11

Searched: 77309 seqs, 7078906 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5\_COMB 2:PCT9\_COMB 3:backfiles1

Statistics: Mean 28.678; Variance 129.097; scale 0.222

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	89	9.1	323	2	PCT-US95-0 Sequence 16, Applicati	1.07e+01
2	89	9.1	323	1	US-08-469- Sequence 16, Applicati	1.07e+01
3	89	9.1	343	1	US-08-279- Sequence 9, Applicati	1.07e+01
4	84	8.6	212	1	US-08-461- Sequence 35, Applicati	2.43e+01
5	84	8.6	748	1	US-08-839- Sequence 2, Applicatio	2.43e+01
6	84	8.6	748	1	US-08-408- Sequence 2, Applicatio	2.43e+01
7	81	8.3	847	1	US-08-781- Sequence 2, Applicatio	3.94e+01
8	81	8.3	847	1	US-08-276- Sequence 2, Applicatio	3.94e+01
9	80	8.2	785	1	US-08-526- Sequence 4, Applicatio	4.62e+01
10	79	8.1	227	1	US-08-414- Sequence 4, Applicatio	5.42e+01
11	79	8.1	357	2	PCT-US94-0 Sequence 12, Applicati	5.42e+01
12	79	8.1	357	1	US-08-145- Sequence 12, Applicati	5.42e+01
13	79	8.1	921	1	US-08-396- Sequence 2, Applicatio	5.42e+01
14	79	8.1	921	1	US-08-818- Sequence 2, Applicatio	5.42e+01
15	78	8.0	119	1	US-08-207- Sequence 20, Applicati	6.35e+01
16	78	8.0	125	1	US-08-313- Sequence 54, Applicati	6.35e+01
17	78	8.0	195	2	PCT-US93-0 Sequence 9, Applicatio	6.35e+01
18	78	8.0	195	1	US-08-063- Sequence 9, Applicatio	6.35e+01
19	78	8.0	677	1	US-08-646- Sequence 13, Applicati	6.35e+01
20	78	8.0	677	1	US-08-188- Sequence 13, Applicati	6.35e+01
21	77	7.9	312	1	US-08-414- Sequence 2, Applicatio	7.44e+01
22	77	7.9	708	1	US-07-797- Sequence 2, Applicatio	7.44e+01
23	77	7.9	708	1	US-08-308- Sequence 2, Applicatio	7.44e+01

24	77	7.9	708	2	PCT-US95-0	Sequence 2, Applicatio	7.44e+01
25	77	7.9	973	1	US-08-162- Sequence 8, Applicatio	7.44e+01	
26	76	7.8	70	1	US-08-476- Sequence 25, Applicati	8.70e+01	
27	76	7.8	70	1	US-08-485- Sequence 25, Applicati	8.70e+01	
28	76	7.8	70	1	US-07-956- Sequence 25, Applicati	8.70e+01	
29	76	7.8	385	1	US-08-036- Sequence 5, Applicatio	8.70e+01	
30	76	7.8	748	1	US-08-369- Sequence 10, Applicati	8.70e+01	
31	76	7.8	748	2	PCT-US95-1 Sequence 10, Applicati	8.70e+01	
32	76	7.8	749	1	US-08-781- Sequence 15, Applicati	8.70e+01	
33	76	7.8	749	1	US-08-276- Sequence 15, Applicati	8.70e+01	
34	76	7.8	905	1	US-08-072- Sequence 2, Applicatio	8.70e+01	
35	76	7.8	3038	1	US-08-450- Sequence 2, Applicatio	8.70e+01	
36	75	7.7	82	1	US-08-461- Sequence 11, Applicati	1.02e+02	
37	75	7.7	82	2	PCT-US93-1 Sequence 11, Applicati	1.02e+02	
38	75	7.7	106	1	US-08-643- Sequence 14, Applicati	1.02e+02	
39	75	7.7	106	1	US-08-335- Sequence 45, Applicati	1.02e+02	
40	75	7.7	106	1	US-08-406- Sequence 14, Applicati	1.02e+02	
41	75	7.7	263	1	US-08-781- Sequence 17, Applicati	1.02e+02	
42	75	7.7	372	1	US-08-278- Sequence 33, Applicati	1.02e+02	
43	75	7.7	372	1	US-08-643- Sequence 33, Applicati	1.02e+02	
44	75	7.7	983	1	US-08-162- Sequence 16, Applicati	1.02e+02	
45	75	7.7	983	1	US-08-167- Sequence 10, Applicati	1.02e+02	

## ALIGNMENTS

RESULT 1  
ID PCT-US95-07289-16 STANDARD; PRT; 323 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE Sequence 16, Application PC/TUS9507289  
CC  
CC Sequence 16, Application PC/TUS9507289  
CC GENERAL INFORMATION:  
CC APPLICANT: Yu, Guo-Liang  
CC APPLICANT: Rosen, Craig  
CC TITLE OF INVENTION: Colon Specific Genes and Proteins  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESS: Carella, Byrne, Bain, Gilfillan, Cecchi,  
CC ADDRESSEE: Stewart & Olstein  
CC STREET: 6 Becker Farm Road  
CC CITY: Roseland  
CC STATE: NJ  
CC COUNTRY: USA  
CC ZIP: 07068-1739

COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent in Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US95/07289  
CC FILING DATE: 06-JUN-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Ferraro, Gregory D.  
CC REGISTRATION NUMBER: 36,134  
CC REFERENCE/DOCKET NUMBER: 325800-265  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 201-994-1700  
CC TELEFAX: 201-994-1744  
CC INFORMATION FOR SEQ ID NO: 16:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 323 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 323 AA; 35941 MW; 553806 CN;

DE Sequence 35, Application US/08461859



CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/408,318  
 CC FILING DATE:  
 CC CLASSIFICATION: 424  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Osman, Richard A  
 CC REGISTRATION NUMBER: 36,627  
 CC REFERENCE/DOCKET NUMBER: A-60845  
 CC TELEPHONE: 415-494-8700  
 CC TELEFAX: 415-494-8771  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 748 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 748 AA; 85940 MW; 3117180 CN;

Query Match 8.6%; Score 84; DB 1; Length 748;  
 Best Local Similarity 23.8%; Pred. No. 2.43e+01;  
 Matches 19; Conservative 27; Mismatches 30; Indels 4; Gaps 4;

Db 522 LTVOSYSDGHLTW-AKCEKHLPKSTFTW-WLEAILDLKKHLPWIDGYVNGFVS 579  
 QY 49 VTLRASFDRTLAWISRGQRKLLSAPFLYPQRFFEVLLFQEGGLKALNGQGLGATS 108  
 Db 580 KEKERL-LLKD-KMPGTFLL 597  
 QY 109 MNQALEQLRELIRISGVQL 128

RESULT 7  
 ID US-08-781-890-2 STANDARD; PRT; 847 AA.  
 AC xxxxxx  
 XX  
 XX  
 DT  
 XX  
 XX  
 DE  
 XX  
 CC Sequence 2, Application US/08781890  
 CC Sequence 2, Application US/08781890  
 CC Patent No. 5710266  
 CC GENERAL INFORMATION:  
 CC APPLICANT: McKnight, Steven L  
 CC APPLICANT: Hou, Jinzhao  
 CC TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
 CC NUMBER OF SEQUENCES: 17  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
 CC STREET: 4 Embarcadero Center, Suite 3400  
 CC CITY: San Francisco  
 CC STATE: California  
 CC COUNTRY: USA  
 CC ZIP: 94111-4187  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/781,890  
 CC FILING DATE: 05-JAN-1997  
 CC CLASSIFICATION: 536  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/276,099  
 CC FILING DATE: 15-JUL-1994  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Osman, Richard Aron

CC REGISTRATION NUMBER: 36,627  
 CC REFERENCE/DOCKET NUMBER: A-59451-1/RAO  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (415) 781-1989  
 CC TELEFAX: (415) 398-3249  
 CC TELEX: 910 277299  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 847 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 847 AA; 94134 MW; 3802900 CN;

Query Match 8.3%; Score 81; DB 1; Length 847;  
 Best Local Similarity 29.6%; Pred. No. 3.94e+01;  
 Matches 21; Conservative 16; Mismatches 26; Indels 8; Gaps 7;

Db 461 AEVGTNRGL-L--PEHF-LFLAQKIFNDNSLSM-EAFQHSVSW-SQFNKEILLGRGFTF 514  
 QY 20 GOVIVYRGVLQEPKHTVSLRDQA-AHAPVTLRASFDRTLAWISRGQRKLLSAPFLF 78  
 Db 515 W-QWFDGVLDL 524  
 QY 79 YPQRFEEVLL 89

RESULT 8  
 ID US-08-276-099A-2 STANDARD; PRT; 847 AA.  
 AC xxxxxx  
 XX  
 XX  
 DT  
 XX  
 XX  
 DE  
 XX  
 CC Sequence 2, Application US/08276099A  
 CC Sequence 2, Application US/08276099A  
 CC Patent No. 5591825  
 CC GENERAL INFORMATION:  
 CC APPLICANT: McKnight, Steven L  
 CC APPLICANT: Hou, Jinzhao  
 CC TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
 CC NUMBER OF SEQUENCES: 17  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
 CC STREET: 4 Embarcadero Center, Suite 3400  
 CC CITY: San Francisco  
 CC STATE: California  
 CC COUNTRY: USA  
 CC ZIP: 94111-4187  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/276,099A  
 CC FILING DATE: 15-JUL-1994  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Osman, Richard Aron  
 CC REGISTRATION NUMBER: 36,627  
 CC REFERENCE/DOCKET NUMBER: A-59451-1/RAO  
 CC TELEPHONE: (415) 781-1989  
 CC TELEFAX: (415) 398-3249  
 CC TELEX: 910 277299  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 847 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear

```

QY 1 MSPRLVPCSHALPQGLSPGV-IIVRGLVLOEPKH 35

RESULT 10
ID US-08-414-625-4 STANDARD; PRT; 227 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE
DE
XX
CC Sequence 4, Application US/08414625
CC Sequence 4, Application US/08414625
CC Patent No. 5563039
CC GENERAL INFORMATION:
CC APPLICANT: Goedel, David V.
CC APPLICANT: Hsu, Hailing
CC TITLE OF INVENTION: INTRACELLULAR SIGNALING PROTEINS AND
CC TITLE OF INVENTION: METHODS OF USE
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/414,625
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-60916/RAO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 494-8700
CC TELEFAX: (415) 494-8771
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 227 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 227 AA; 25323 MW; 222467 CN;

Query Match 8.1%; Score 79; DB 1; Length 227;
Best Local Similarity 41.9%; Pred. No. 5.42e+01;
Matches 18; Conservative 11; Mismatches 9; Indels 5; Gaps 5

Db 3 LOAVREGALRTALQ-RCM-APALAQEALRLQL-ELR-AGAEQL 41
QY 87 LLLFQEGGKLALNGQGLGATSMNQALE-OLRELIRISGVQL 128

RESULT 11
ID PCT-US94-00545-12 STANDARD; PRT; 357 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE
DE
XX
CC Sequence 12, Application PC/TUS9400545
CC Sequence 12, Application PC/TUS9400545
CC GENERAL INFORMATION:
CC APPLICANT: Rao, Anjana

```

CC APPLICANT: Hogan, Patrick Gerald  
CC APPLICANT: McCaffrey, Patricia  
CC APPLICANT: Jain, Jugnu  
CC TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE  
CC TITLE OF INVENTION: DNA-BINDING PROTEIN  
CC NUMBER OF SEQUENCES: 22  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fish & Richardson  
CC STREET: 225 Franklin Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
CC COMPUTER: IBM PS/2 Model 502 or 55SX  
CC OPERATING SYSTEM: MS-DOS (Version 5.0)  
CC SOFTWARE: Wordperfect (Version 5.1)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US94/00545  
CC FILING DATE: 18-JAN-1994  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/145,006  
CC FILING DATE: October 29, 1993  
CC APPLICATION NUMBER: 08/017,052  
CC FILING DATE: February 11, 1993  
CC APPLICATION NUMBER: 08/006,067  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fraser, Janis K.  
CC REGISTRATION NUMBER: 34,819  
CC REFERENCE/DOCKET NUMBER: 04590/007001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 542-5070  
CC TELEFAX: (617) 542-8906  
CC TELEX: 200154  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 357  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 357 AA; 37623 MW; 700002 CN;  
SQ  
Query Match 8.1%; Score 79; DB 2; Length 357;  
Best Local Similarity 36.4%; Pred. No. 5.42e+01;  
Matches 12; Conservative 12; Mismatches 8; Indels 1; Gaps 1;  
Db 51 LSPRIETPSHELIVAGVPLMRDA-GLIVEQP 82  
:||||: ||| |:::| :|:::|  
QY 1 MSPRLVPCSHALPQGLSPGVIVRGLVLEP 33  
RESULT 12  
ID US-08-145-006C-12 STANDARD; PRT; 357 AA.  
XX AC xxxxxx  
XX AC xxxxxx  
DT DT  
XX XX  
DE DE  
XX XX  
Sequence 12, Application US/08145006C  
Sequence 12, Application US/08145006C  
Patent No. 5656452  
GENERAL INFORMATION:  
CC APPLICANT: Rao, Anjana  
CC APPLICANT: Hogan, Patrick Gerald  
CC APPLICANT: McCaffrey, Patricia  
CC APPLICANT: Jain, Jugnu  
CC TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE  
CC TITLE OF INVENTION: DNA-BINDING PROTEIN  
CC NUMBER OF SEQUENCES: 21

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fish & Richardson P.C.  
CC STREET: 225 Franklin Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
CC COMPUTER: IBM PS/2 Model 502 or 55SX  
CC OPERATING SYSTEM: MS-DOS (Version 5.0)  
CC SOFTWARE: Wordperfect (Version 5.1)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/145,006C  
CC FILING DATE: October 29, 1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/017,052  
CC FILING DATE: February 11, 1993  
CC APPLICATION NUMBER: 08/006,067  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fraser, Janis K.  
CC REGISTRATION NUMBER: 34,819  
CC REFERENCE/DOCKET NUMBER: 04590/007001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 542-5070  
CC TELEFAX: (617) 542-8906  
CC TELEX: 200154  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 357  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 357 AA; 37623 MW; 700002 CN;  
SQ  
Query Match 8.1%; Score 79; DB 1; Length 357;  
Best Local Similarity 36.4%; Pred. No. 5.42e+01;  
Matches 12; Conservative 12; Mismatches 8; Indels 1; Gaps 1;  
Db 51 LSPRIETPSHELIVAGVPLMRDA-GLIVEQP 82  
:||||: ||| |:::| :|:::|  
QY 1 MSPRLVPCSHALPQGLSPGVIVRGLVLEP 33  
RESULT 13  
ID US-08-396-479B-2 STANDARD; PRT; 921 AA.  
XX AC xxxxxx  
XX AC xxxxxx  
DT DT  
XX XX  
DE DE  
XX XX  
Sequence 2, Application US/08396479B  
Sequence 2, Application US/08396479B  
Patent No. 5612455  
GENERAL INFORMATION:  
CC APPLICANT: HOEY, Timothy  
CC TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
CC NUMBER OF SEQUENCES: 18  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
CC STREET: 4 Embarcadero Center, Suite 3400  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30

4:

Db 67 RVTMLRDTSKNOFSLRLSSVTAADTAIYICARHGDDP-AWFAYWGOGSLVTVS 118  
Qy 26 RGLVLOE-PK-HFTVSLRD-QAAHAPVTILRASFADETLAWISRWGOKKLISAP 75

Search completed: Tue Dec 15 15:56:35 1998  
Job time : 8 secs.

(T)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```
Run on: Tue Dec 15 15:54:30 1998; MasPar time 7.48 Seconds
        606.724 Million cell updates/sec
Tabular output not generated.
```

Sequence: 1 MSPRLEVPCSHALPQGLSPG.....LEQLRELRIKSGSVQLYCVHS 133

Searched: 107076 seqs, 34141958 residues

```
Database:
pir56
  1:pir1 2:pir2 3:pir3 4:pir4
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result	Query	Score	Match	Length	DB	ID	Description	Pred. No.
--------	-------	-------	-------	--------	----	----	-------------	-----------

1	174	17.8	145	2	A55932	galactin-5 - rat	7.78e-12
2	130	14.3	317	2	JC6147	prostate carcinoma tu	1.19e-06
3	136	13.9	326	2	I55469	galactin-7 - human	4.56e-06
4	112	11.4	279	2	A42846	beta-galactoside-bind	9.92e-03
5	109	11.1	316	2	A55975	galactin-8 - rat	2.46e-02
6	109	11.1	324	2	A46631	lactose-binding lecti	2.46e-02
7	106	10.8	448	2	A27522	UDP-N-acetylglucosami	6.03e-02
8	99	10.1	135	1	LNCH14	14K beta-galactoside-	4.60e-01
9	98	10.0	308	2	SI3584	nosf protein - pseudo	4.60e-01
10	95	9.7	746	2	A40636	enterobactin receptor	1.41e+00
11	95	9.7	1321	2	A60165	sodium channel protei	1.41e+00
12	93	9.5	493	2	A65126	probable general secr	2.45e+00
13	93	9.5	1522	2	JC1101	sodium channel protei	2.45e+00
14	92	9.4	583	2	A29154	complement factor I (	3.22e+00
15	92	9.4	620	1	Q6ECSR	glutathione-regulated	3.22e+00
16	91	9.3	703	2	B41538	ATP-binding cassette	4.21e+00
17	90	9.2	242	2	A49800	galactoside-binding p	5.50e+00
18	90	9.2	510	2	A64101	multidrug resistance	5.50e+00
19	89	9.2	761	2	S20458	pqgf protein - Klebsi	5.50e+00
20	89	9.1	225	2	A48640	global nitrogen regul	7.18e+00
21	89	9.1	833	2	JQ1866	hypothetical 87.1K pr	7.18e+00
22	89	9.1	862	2	S57964	lipoxygenase (EC 1.13	7.18e+00
23	88	9.0	137	2	B40361	virC-region hypothei	9.34e+00

```

Db 70 NTQINNSGPEERSLPGSMFSGRQRF-SYWLCEGHCFAVDGQHICEVSHRLMLPD 128
QY 63 ---ISR-WG--OKKLIAPFLFY-PQRFVEVLLFQEGGLKALNGQGLGATSMNQALEQ 116

Db 129 INTLEVAGDIQL 140
QY 117 LRELRISSGVQL 128

RESULT 2
ENTRY #type complete
TITLE prostate carcinoma tumor antigen 1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Sep-1997 #sequence_revision 03-Sep-1997 #text_change
31-Oct-1997
ACCESSIONS JC6147
REFERENCE JC6147
#authors Su, Z.Z.; Lin, J.; Shen, R.; Fisher, P.E.; Goldstein, N.I.;
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:7252-7257
#title Surface-epitope masking and expression cloning identifies the
human prostate carcinoma tumor antigen gene PCTA-1 a member
of the galectin gene family.
#accession JC6147
#molecule_type mRNA
#residues 1-317 #label SUA
#cross-references GB:L78132; NID:g1932711; PID:g1932712
COMMENT This protein is a therapeutic reagent for intervention in pervasive
and fatal neoplastic disease. It is involved in an earlier
genetic change in human prostate cancer development. This protein
is a member of the galectin family.
KEYWORDS tumor
SUMMARY #length 317 #molecular-weight 35719 #checksum 1508

Query Match 14.3%; Score 140; DB 2; Length 317;
Best Local Similarity 24.6%; Pred. No. 1.19e-06;
Matches 34; Conservative 41; Mismatches 56; Indels 7; Gaps 5;

Db 180 TPOLSLPFAARLNTMPGRTVVQGEVNAKSFNVDDLAKGSKDIALHNLPLNIKAF 239
QY 2 SPLEVPCHSHALPQGLSPGVQVIIVRGLVQEPKHFTVSL-RDQAAHAPVTLRASFAQRTL 60

Db 240 VRNSFTQESWGEERNITSPFPFGMYFEMIYCDVREFKAVNGVHSLYKHFKE-LS 298
QY 61 AWIS--R--WGKKLIAPFLFY-PQRFVEVLLFQEGGLKALNG-QGLGATSMNQALE 115

Db 299 SIDTLEINGDIHLEVR 316
QY 116 QLRELRISSGVQLYCVHS 133

RESULT 3
ENTRY #type complete
TITLE galectin-7 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-May-1996 #sequence_revision 24-May-1996 #text_change
24-May-1996
ACCESSIONS I55469
REFERENCE I55469
#authors Madsen, P.; Rasmussen, H.H.; Flint, T.; Gromov, P.; Kruse,
T.A.; Honore, B.; Vorum, H.; Celis, J.E.
#journal J. Biol. Chem. (1995) 270:5823-5829
#title Cloning, expression, and chromosome mapping of human
galectin-7.
#cross-references MUID:95197604
#accession I55469
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-136 #label RES
#cross-references GB:L07769; NID:g182131; PID:g182132
SUMMARY #length 136 #molecular-weight 15075 #checksum 8853

```

```

Query Match 13.9%; Score 136; DB 2; Length 136;
Best Local Similarity 28.1%; Pred. No. 4.56e-06;
Matches 38; Conservative 39; Mismatches 47; Indels 11; Gaps 6;

Db 3 NVPHKSLPEGIRGTVLIRGLVPPNARFHVNLGCEQGSDAALHFNRLDTSEVVF 62
QY 6 EVPCSHALPQGLSPGVQVIIVRGLVQEPKHFTVSL--RDQAAHAPVTLRASFAQRTLAW 62

Db 63 NSKEQSGWREERGPGVYFQR-QQPF-EVLLIADDDGFK-AVVGDAQYHHFHRPLARV 119
QY 63 ISR----WG--OKKLIAPFLFY-PQRFVEVLLFQEGGLKALNGQGLGATSMNQALEQ 117

Db 120 RLVEVGDDVOLDVSR 134
QY 118 RELRISSGVQLYCVH 132

RESULT 4
ENTRY #type complete
TITLE beta-galactoside-binding protein GBP - Caenorhabditis elegans
ALTERNATE_NAMES beta-galactoside-binding lectin homolog, 32K
ORGANISM #formal_name Caenorhabditis elegans
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
09-Sep-1997
ACCESSIONS A42846; S27798
REFERENCE A42846
#authors Hirabayashi, J.; Satoh, M.; Kasai, K.
#journal J. Biol. Chem. (1992) 267:15485-15490
#title Evidence that Caenorhabditis elegans 32-kDa
beta-galactoside-binding protein is homologous to
vertebrate beta-galactoside-binding lectins. cDNA cloning
and deduced amino acid sequence.
#cross-references MUID:92348399
#accession A42846
#status preliminary
#molecule_type mRNA
#residues 1-279 #label HIR
#cross-references EMBL:M94671; NID:g156209; PID:g156210
#note sequence extracted from NCBI backbone (NCBIN:109554)
SUMMARY #length 279 #molecular-weight 31809 #checksum 854

Query Match 11.4%; Score 112; DB 2; Length 279;
Best Local Similarity 35.9%; Pred. No. 9.92e-03;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 11 VPIRSVLQKEPQGTIVKGTIDESORFTINLHSKTA 49
QY 7 VPCSHALPQGLSPGVQVIIVRGLVQEPKHFTVSLRDQAA 45

RESULT 5
ENTRY #type complete
TITLE galectin-8 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change
10-Sep-1997
ACCESSIONS A55975
REFERENCE A55975
#authors Hadari, Y.R.; Paz, K.; Dekel, R.; Mestrovic, T.; Accili, D.;
Zick, Y.
#journal J. Biol. Chem. (1995) 270:3447-3453
#title Galectin-8. A new rat lectin, related to galectin-4.
#accession A55975
#status preliminary
#molecule_type mRNA
#residues 1-316 #label HAD
#cross-references GB:U09824; NID:g171031; PID:g171032
SUMMARY #length 316 #molecular-weight 36038 #checksum 9989

Query Match 11.1%; Score 109; DB 2; Length 316;
Best Local Similarity 22.1%; Pred. No. 2.46e-02;
Matches 30; Conservative 41; Mismatches 58; Indels 7; Gaps 5;

```

Db	191	HLSLPEARLNAMSGPRTVVVKGENVNTATSNVDLVAGSRDIALHLNPLNKAFAVR	240
Qy	4	RLEVPCHALPOGSLGVIIIVRGVLQEPKHFTVSL-RDQAAHAPVTLRASFAADRTLAW	62
Db	241	NSFLQDAMGEERNITCFPFSSGGMFEMIIYCDVREFKVAVNGVHSLEYKHREFKD-LSSI	299
Qy	63	IS-R-WGQKKLISAPFFLYPQRFVEVLLFQGGGLKALNG-OGLGATSMNQALEQL	117
Db	300	DTLAVDGDIRLLDVR	315
Qy	118	RELRTSGSVQLYCVHS	133
RESULT	6		
ENTRY		A46631	#type complete
TITLE			lactose-binding lectin L-36 - rat
ALTERNATE_NAMES			galectin-4
ORGANISM			#formal_name Rattus norvegicus #common_name Norway rat
DATE			21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
ACCESSIONS		A46631; S69096	
REFERENCE		A46631	
#authors			Oda, Y.; Herrmann, J.; Gitt, M.A.; Turck, C.W.; Burlingame, A.L.; Barondes, S.H.; Leffler, H.
#journal			J. Biol. Chem. (1993) 268:5929-5939
#title			Soluble lactose-binding lectin from rat intestine with two different carbohydrate-binding domains in the same peptide chain.
#cross-references		MUID:93194902	
#accession		A46631	
#status			preliminary
#molecule_type			nucleic acid
#residues			1-324 #label ODA
#cross-references		GB:M73553; NID:g294571; PID:g294572	
#experimental_source			intestine
#note			sequence extracted from NCBI backbone (NCBIN:128409, NCBI:P:128410)
REFERENCE		S69096	
#authors			Tardy, F.; Deviller, P.; Louisot, P.; Martin, A.
#journal			FEBS Lett. (1995) 359:169-172
#title			Purification and characterization of the N-terminal domain of galectin-4 from rat small intestine.
#accession		S69096	
#molecule_type			protein
#residues			13-37;44-50,'E',52-66 #label TAR
KEYWORDS			lectin
SUMMARY			#length 324 #molecular-weight 36347 #checksum 2058
Query Match			11.1%; Score 109; DB 2; Length 324;
Best Local Similarity			23.7%; Pred. No. 2.46e-02;
Matches			31; Conservative 38; Mismatches 56; Indels 6; Gaps 5;
Db	194	VPYVGTQGLGTLARRPIILKGVYLPATAKLLINFKVSGTGDIAFHMPRICDCVVRNSYM	253
Qy	7	VPCSHALPOGSLGVIIIVRGVLQEPKHFTVSLR-DQAAHAPVTLRASFAADRTL--AWI	63
Db	254	NGSWGSEE-RKIPYNPFGAGQFDLSIRCTGDRFKVPANGQHILFDFSHRFOAFQVRDML	312
Qy	64	S-RWGQKKLISAPF-LFYQRFVEVLLFQGGGLKALNGQGLGATSMNQALEQLREL	121
Db	313	IKGDTLTSVQ	323
Qy	122	ISGVSQLYCVH	132
RESULT	7		
ENTRY		A27522	#type complete
TITLE			UDP-N-acetylglucosamine--dolichyl-phosphate
ALTERNATE_NAMES			N-acetylglucosaminophosphotransferase (EC 2.7.8.15) - yeast (Saccharomyces cerevisiae)
ORGANISM			protein YBR1628; protein YBR243C
DATE			#formal_name Saccharomyces cerevisiae
			15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change

Query Match 9.7%; Score 95; DB 2; Length 1321;

#domain	transmembrane	#status	predicted	#label	TM3\
#domain	transmembrane	#status	predicted	#label	TM4\
#domain	transmembrane	#status	predicted	#label	TM5\
#domain	transmembrane	#status	predicted	#label	TM6\
#domain	transmembrane	#status	predicted	#label	TM7\
#domain	transmembrane	#status	predicted	#label	TM8\
#domain	transmembrane	#status	predicted	#label	TM9\
#domain	transmembrane	#status	predicted	#label	TM10\
#domain	transmembrane	#status	predicted	#label	TM11\
#domain	transmembrane	#status	predicted	#label	TM12\
#domain	transmembrane	#status	predicted	#label	TM13\
#domain	transmembrane	#status	predicted	#label	TM14\
#domain	transmembrane	#status	predicted	#label	TM15\
#domain	transmembrane	#status	predicted	#label	TM16\
#domain	transmembrane	#status	predicted	#label	TM17\
#domain	transmembrane	#status	predicted	#label	TM18\
#domain	transmembrane	#status	predicted	#label	TM19\
#domain	transmembrane	#status	predicted	#label	TM20\
#domain	transmembrane	#status	predicted	#label	TM21\
#domain	transmembrane	#status	predicted	#label	TM22\
#domain	transmembrane	#status	predicted	#label	TM23\
#domain	transmembrane	#status	predicted	#label	TM24\
#length 1522 #molecular-weight 17413 #checksum 2729					
Query Match                9.5%; Score 93; DB 2; Length 1522;					
Best Local Similarity 35.0%; Pred. No. 2.45e+00;					
Matches	14;	Conservative	12;	Mismatches 14; Indels 0; Gaps 0;	
Db	878	SRLTLRALPLRAVSQGHKIIVNALMLSPISFNVLV	917		
QY	50	TLRASFADRTLAWISRWGKKLISAPFLYPQRFVEILL	89		
RESULT	14				
ENTRY	A29154	#type complete			
TITLE	Complement factor I (EC 3.4.21.45) precursor - human				
ALTERNATE_NAMES	C3b/C4b inactivator				
ORGANISM	#formal_name Homo sapiens	#common_name man			
DATE	31-Dec-1988	#sequence_revision 31-Dec-1988	#text_change		
	10-Oct-1997.				
ACCESSIONS	A29154; A28434; S66420				
REFERENCE	A29154				
#authors	Catterall, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.				
#journal	Biochem. J. (1987) 242:849-856				
#title	Characterization of the primary amino acid sequence of human complement control protein factor I from an analysis of cDNA clones.				
#cross-references	MUID:87241401				
#accession	A29154	#molecule_type mRNA			
#residues	1-583	#label CAT			
#cross-references	GB:Y00318				
REFERENCE	A28434				
#authors	Goldberger, G.; Bruns, G.A.P.; Rits, M.; Edge, M.D.; Kwiatkowski, D.J.				
#journal	J. Biol. Chem. (1987) 262:10065-10071				
#title	Human complement factor I: Analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4.				
#cross-references	MUID:87280021				
#accession	A28434	#molecule_type mRNA			
#residues	1-557,'F',559-583	#label GOL			
#cross-references	GB:J02770; NID:g182606; PID:g182607				
REFERENCE	S66420				
#authors	Ullman, C.G.; Haris, P.I.; Smith, K.F.; Sim, R.B.; Emery, V.C.; Perkins, S.J.				
#journal	FEBS Lett. (1995) 371:199-203				
#title	beta-Sheet secondary structure of an LDL receptor domain from complement factor I by consensus structure predictions and spectroscopy.				
#accession	S66420	#molecule type protein			

```

#residues      258-269 #label ULL
GENETICS
#gene          GDB:IF
#cross-references GDB:120077; OMIM:217030
#map_position  4q24-q25
CLASSIFICATION #superfamily LDL receptor ligand-binding repeat homology;
               #trypsin homology
KEYWORDS       glycoprotein; hydrolase; serine proteinase
FEATURE
1-18           #domain signal sequence #status predicted #label SIG\
19-335        #product complement factor I heavy chain #status
               #predicted #label CFH\
221-255        #domain LDL receptor ligand-binding repeat homology
               #label LDL1\
259-293        #domain LDL receptor ligand-binding repeat homology
               #label LDL2\
340-583        #product complement factor I light chain #status
               #predicted #label CFL\
340-569        #domain trypsin homology #label TRY\
70,103,177,464,494, #binding_site carbohydrate (Asn) (covalent) #status
536            predicted
SUMMARY        #length 583 #molecular-weight 65720 #checksum 3015

Query Match      9.4%; Score 92; DB 2; Length 583;
Best Local Similarity 52.4%; Pred. No. 3.22e+00;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 485 QWGEVKLISCKFYGNRFYE 505
QY 65 RWGQKKLISAPFLFPQRF 85

RESULT 15
ENTRY   QOECRD #type complete
TITLE   glutathione-regulated potassium efflux system protein kefc -
        Escherichia coli
ORGANISM
        #formal_name Escherichia coli
        #date 05-Apr-1983 #sequence_revision 01-Mar-1996 #text_change
        14-Nov-1997
ACCESSIONS
REFERENCE S40568; S15165; A04444; G64725
#authors   Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.;
            Fujita, N.; Isono, K.; Mizobuchi, K.; Nakata, A.
#submission submitted to the EMBL Data Library, December 1992
#description Systematic sequencing of the Escherichia coli genome:
            analysis of the 0-2.4min region.
#accession S40568
#status    preliminary
#molecule_type DNA
#residues  1-620 #label YUR
#cross-references EMBL:DL0483; NID:g216434; PID:d1001795; PID:g216472
REFERENCE S15165
#authors   Munro, A.W.; Ritchie, G.Y.; Lamb, A.J.; Douglas, R.M.; Booth,
            I.R.
#journal   Mol. Microbiol. (1991) 5:607-616
#title     The cloning and DNA sequence of the gene for the
            glutathione-regulated potassium-efflux system kefc of
            Escherichia coli.
#cross-references MUID:9126044
#accession S15165
#molecule_type DNA
#residues  1-620 #label MUN
#cross-references EMBL:X56742; NID:g41874; PID:g41875
REFERENCE A93704
#authors   Smith, D.R.; Calvo, J.M.
#journal   Nucleic Acids Res. (1980) 8:2255-2274
#title     Nucleotide sequence of the Escherichia coli gene coding for
            dihydrofolate reductase.
#cross-references MUID:81053692
#accession A04444
#molecule_type DNA
#residues  500-620 #label SMI

```

```

#experimental_source strain K12
REFERENCE      A64720
#authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal      Science (1997) 277:1453-1462
#title        The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession    G64725
#status       preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues     1-620 #label BLAT
#cross-references GB:AE000115; GB:U000096; NID:gl786230; PID:gl786232;
            UMGp:b0047
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene         kefc
#map_position 1 min
CLASSIFICATION #superfamily glutathione-regulated potassium efflux system
               protein kefc
KEYWORDS       potassium transport; transmembrane protein
FEATURE        5-26
               #domain transmembrane #status predicted #label TM1\
48-70          #domain transmembrane #status predicted #label TM2\
77-96          #domain transmembrane #status predicted #label TM3\
103-124        #domain transmembrane #status predicted #label TM4\
130-144        #domain transmembrane #status predicted #label TM5\
149-170        #domain transmembrane #status predicted #label TM6\
SUMMARY        #length 620 #molecular-weight 67795 #checksum 3086

Query Match      9.4%; Score 92; DB 1; Length 620;
Best Local Similarity 34.9%; Pred. No. 3.22e+00;
Matches 30; Conservative 19; Mismatches 27; Indels 10; Gaps 8;

Db 181 FALSALKVAGALVVLGLRYTRPALRFVARSGLREVSVALFLV--FGFG-LLL-EE 236
QY 36 FTVS-LRDOAAHAPVTLRASFAADR-TLAWISRWGQKKLISAPFLFPQRF-FEVLLLFQE 92

Db 237 VGLSMANGAFLAGVLLASSEYRHALE 262
QY 93 GGLKLALNG--QG-LGATSMNQALE 115

Search completed: Tue Dec 15 15:54:53 1998
Job time : 23 secs.

```

\*\*\*\*\*

WIRE (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 15:55:45 1998; MasPar time 10.89 Seconds  
Tabular output not generated. 608.412 Million cell updates/sec

Title: >US-09-109-864-2  
Description: (1-133) from US09109864.pep  
Perfect Score: 980  
Sequence: 1 MSPRLEVPCHALPQGLSPG.....LEQLRELRISGVOLYCVHS 133

Scoring table: PAM 150  
Gap 11

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb16  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 42.859; Variance 78.955; scale 0.543

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	146	14.9	354	11	36 KD BETA-GALACTOSIDE	6.48e-09
2	141	14.4	318	4	GALECTIN-8.	4.56e-08
3	113	11.5	136	11	GALECTIN-7.	1.46e-03
4	112	11.4	285	5	W09H1.6B.	2.07e-03
5	111	11.3	262	13	GALECTIN-3.	2.93e-03
6	111	11.3	301	11	GALECTIN-6.	2.93e-03
7	110	11.2	283	5	GALECTIN-7.	4.14e-03
8	104	10.6	1262	5	F52H3.7.	3.18e-02
9	102	10.4	280	5	BETA-GALACTOSIDE-BINDI	6.19e-02
10	101	10.3	132	11	GALECTIN-7.	8.61e-02
11	100	10.2	195	14	BETA-A PROTEIN.	1.19e-01
12	98	10.0	573	10	TERPENE CYCLASE LIKE P	2.28e-01
13	98	10.0	603	11	COMPLEMENT COMPONENT F	2.28e-01
14	97	9.9	1309	10	REVERSE TRANSCRIPTASE.	3.15e-01
15	96	9.8	295	2	PUTATIVE INNER MEMBRAN	4.33e-01
16	95	9.7	703	6	ABC-TRANSPORTER.	5.95e-01
17	95	9.7	1320	5	SODIUM CHANNEL PROTEIN	5.95e-01
18	94	9.6	278	5	GALECTIN.	8.15e-01
19	92	9.4	166	2	HYPOTHETICAL 19.2 KD P	1.52e+00
20	92	9.4	278	5	GALECTIN.	1.52e+00

21	92	9.4	703	6	Q28435	ABC-TRANSPORTER.	1.52e+00
22	92	9.4	703	6	Q28437	ABC-TRANSPORTER.	1.52e+00
23	92	9.4	703	6	Q28436	ABC-TRANSPORTER.	1.52e+00
24	92	9.4	2124	2	Q53840	POLYKETIDE SYNTHASE (F	2.06e+00
25	91	9.3	260	5	O17270	R148.7 PROTEIN.	2.80e+00
26	90	9.2	324	2	O07208	HYPOTHETICAL 35.0 KD P	2.80e+00
27	90	9.2	928	14	P88954	ORF 63.	2.80e+00
28	89	9.1	64	5	Q29369	30 KDA S-TYPE LECTIN (	3.79e+00
29	89	9.1	399	2	Q56257	HEMOLYSIN.	3.79e+00
30	89	9.1	862	10	Q43800	LIPOXYGENASE (EC 1.13.	3.79e+00
31	88	9.0	137	2	Q56973	KIM5.	5.11e+00
32	88	9.0	291	10	P93740	HYPOTHETICAL 32.4 KD P	5.11e+00
33	88	9.0	295	2	Q05769	HYPOTHETICAL 32.3 KD P	5.11e+00
34	87	8.9	198	14	Q89433	COAT PROTEIN.	6.88e+00
35	87	8.9	198	14	Q65721	COAT PROTEIN.	6.88e+00
36	87	8.9	346	1	O27843	CONSERVED PROTEIN.	6.88e+00
37	87	8.9	424	1	O26646	CATIONIC AMINO ACID TR	6.88e+00
38	87	8.9	445	1	O59214	445AA LONG HYPOTHETICA	6.88e+00
39	87	8.9	1670	5	Q23901	HISTIDINE KINASE.	6.88e+00
40	87	8.9	1711	2	P96311	ENDOGLUCANASE A (EC 3.	6.88e+00
41	86	8.8	309	2	Q47455	PLASMOD PRJ1004 DNA.	9.24e+00
42	86	8.8	353	8	O47472	CYTOCHROME OXIDASE SUB	9.24e+00
43	86	8.8	403	1	O58825	403AA LONG HYPOTHETICA	9.24e+00
44	86	8.8	642	10	O64397	RECEPTOR-LIKE PROTEIN	9.24e+00
45	86	8.8	682	2	O67400	HYPOTHETICAL 77.8 KD P	9.24e+00

ALIGNMENTS

RESULT 1  
ID O35866 PRELIMINARY; PRT; 354 AA.  
AC O35866;  
DT 01-JAN-1998 (TREMREL. 05, CREATED)  
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
DE 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE 36 KD BETA-GALACTOSIDE BINDING LECTIN.  
OS RATUUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-SMALL INTESTINE;  
RA WADA J., OTA K., KOMAR A., KANWAR Y.S.;  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U72741; G2351553;  
DR PROSITE; P500309; LECTIN GALACTOSIDE; 2.  
DR PFAM; PF00337; Gal-bind\_lectin.  
KW LECTIN.  
SQ SEQUENCE 354 AA; 39946 MW; C7E7FEA CRC32;

Query Match 14.9%; Score 146; DB 11; Length 354;  
Best Local Similarity 23.4%; Pred. No. 6.48e-09;  
Matches 32; Conservative 42; Mismatches 57; Indels 6; Gaps 4;  
Db 219 TPAYTIPFTSIPNGYPKSNISGVLPDAKRFHINLRC-GGDIAFHLPNRFNEKVVV 277  
QY 2 SPRLVPCSHALPQGLSPGVIIIVRGVLQPKHFTVSLRQAAHAPVTLRSFADRTLA 61  
Db 278 RNTQINNSGWPERSLPGRMFNRGQSFVWILCEGHCFFKVDGQHCICYYHRLNLPD 337  
QY 62 W---ISR-WG-QKKLISAPFLYPQRFVLLLFQEGGLKALNGGLGATSMNQALEQ 116  
Db 338 INTLEVAGDIQIHTVQT 354  
QY 117 LRELIRISGVOLYCVHS 133

RESULT 2  
ID O15215 PRELIMINARY; PRT; 318 AA.  
AC O15215;  
DT 01-JAN-1998 (TREMREL. 05, CREATED)  
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)

```
DE GALECTIN-8.
GN GAL-8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN HIPPOCAMPUS;
RA HADARI Y.R.; EISENSTEIN M.; ZAKUT R.; ZICK Y.;
RL TRENDS GLYCOSCI. GLYCOTECHNOL. 9:103-112(1997).
DR EMBL; X91790; E199407; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
DR PFAM; PF00337; Gal-bind_lectin.
DR SEQUENCE 318 AA; 35864 MW; 8FD554C CRC32;

Query Match 14.4%; Score 141; DB 4; Length 318;
Best Local Similarity 24.6%; Pred. No. 4.56e-08;
Matches 34; Conservative 41; Mismatches 56; Indels 7; Gaps 5;

Db 181 TQRLRPFARLNPMPGRTVVVKGEVNAKSFNVDDLAKGSKDIALHNLPLNLIKAF 240
QY 2 SPRLVPCSHALPQGLSPGQVIIVRGLVQEPKHFVSL-RDQAAHAPVTLRASFAADRTL 60
Db 241 VNSFLQESWGGEERNITSPFSPGMYEEMIIYCDVREKRVAVNGVHSLEYKHFE-LS 299
QY 61 AWIS--R--WGKKLSAPFLFPQRFEEVLLLFQEGGLKALNG--OGLGATSMNQOALE 115
Db 300 SIDLTEINGDIHLLEVR 317
QY 116 QLRELRISSVQLCVHS 133

RESULT 3
ID O54974 PRELIMINARY; PRT; 136 AA.
AC O54974;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE GALECTIN-7.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C59 BLACKC/6;
RX MEDLINE; 95246905.
RA MAGNALDO T.; BERNERD F.; DARMON M.;
RL DEV. BIOL. 168:259-271(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C59 BLACKC/6;
RA MAGNALDO T.; FOWLIS D.; DARMON M.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF038562; G2708654; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
SQ SEQUENCE 136 AA; 15173 MW; 61A0164B CRC32;

Query Match 11.5%; Score 113; DB 11; Length 136;
Best Local Similarity 27.7%; Pred. No. 1.46e-03;
Matches 36; Conservative 37; Mismatches 44; Indels 13; Gaps 9;

Db 9 SLPGQVRVTVNRIRGLVPDQAGRFHVNILCGEEQADAAALHFNPRLDTSSEVVNTKQOG 68
QY 12 ALPGQLSPQVIIVRGLVQEPKHFVSL--RDQAAHAPVTLRASFAADRTLAW-I---S 64
Db 69 KNGREERGTFQPR-GQPF-EVLLIATEEGFK-AVVGDD-EYLHFHRLPPARVRLVEY 124
QY 65 RRG-OKKLISAPFLFPQRFEEVLLLFQEGGLKALNGOGLGATSMNQO-ALQOLRELI 122
Db 125 GGDVQLHSLN 134
QY 123 SGSVQLCVH 132

RESULT 4
ID O45904 PRELIMINARY; PRT; 285 AA.
AC O45904;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE W09H1.6B.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA SMYE R.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R.; AINSCOUGH R.; ANDERSON K.; BAYNES C.; BERKS M.; BONFIELD J.;
RA BURTON J.; CONNELL M.; COPSEY T.; COOPER J.; COULSON A.; CRAXTON M.;
RA DEAR S.; DU Z.; DURBIN R.; FAVELLO A.; FUTON L.; GARDNER A.; GREEN P.;
RA HAWKINS T.; HILLIER L.; JIER M.; JOHNSTON L.; JONES M.; KERSHAW J.;
RA KIRSTEN J.; LAISTER N.; LATREILLE P.; LIGHTNING J.; LLOYD C.;
RA KIMMURAY A.; MORTIMORE B.; O'CALLAGHAN M.; PARSONS J.; PERCY C.;
RA RIFKEN L.; ROOPRA A.; SAUNDERS D.; SHOWNKEEN R.; SMALDON N.; SMITH A.;
RA SONNHAMMER E.; STADEN R.; SULTON J.; THIERRY-MIEG J.; THOMAS K.;
RA VAUDIN M.; VAUGHAN K.; WATERSTON R.; WATSON A.; WEINSTOCK L.;
RA WILKINSON-SPROAT J.; WOHLDMAN P.;
RL NATURE 368:32-38(1994).
DR EMBL; Z82081; E1247370; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 2.
SQ SEQUENCE 285 AA; 32683 MW; 920E5744 CRC32;

Query Match 11.4%; Score 112; DB 5; Length 285;
Best Local Similarity 35.9%; Pred. No. 2.07e-03;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 17 VPYRSVLOEKEFECQTLIVKGSIDESQRTINLHSKTA 55
QY 7 VPCSHALPQGLSPGQVIIVRGLVQEPKHFVSLRDQAA 45

RESULT 5
ID Q90713 PRELIMINARY; PRT; 262 AA.
AC Q90713;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GALECTIN-3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RA NURMINSKAYA M.V.; LINSSENAYER T.F.;
RL DEV. DYN. 0:0-0(0).
DR EMBL; U50339; G1389600; -.
DR PFAM; PF00337; Gal-bind_lectin.
KW LECTIN.
SQ SEQUENCE 262 AA; 28155 MW; FD6F1E17 CRC32;

Query Match 11.3%; Score 111; DB 13; Length 262;
Best Local Similarity 25.0%; Pred. No. 2.93e-03;
Matches 34; Conservative 30; Mismatches 56; Indels 6; Gaps 5;

Db 122 AAPKVPYDPLPAGLMPRLIITITGVNSNPFRSDFKRGODIAFHFRPFREDHKRV 181
QY 2 SPRLVPCSHALPQGLSPGQVIIVRGLVQEPKHFVSL-RDQAA-A-HAPVTLRASFA 58
Db 162 IVCSNMFQNNWGREERTAPRFFPGTFKQLVLCGDHFKVAVNDALHLLQFNFEKKLN 241
QY 59 TL-AWI--SRWGKKLISAPFLFPQRFEEVLLLFQEGGLKALNGOGLGATSMNQOALE 115
Db 242 GITKLCIAGDITLTSV 257
```

Db 116 QRLRLRSGSVQLVCV 131  
 QY 116 QRLRLRSGSVQLVCV 131  
 RESULT 6  
 ID O54891 PRELIMINARY; PRT; 301 AA.  
 AC O54891  
 DT 01-JUN-1998 (TREMUREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMUREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMUREL. 06, LAST SEQUENCE UPDATE)  
 DE GALECTIN-6.  
 GN LGALS6.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA GITT M.A., COLNOT C., POIRIER F., NANI K.J., BARONDES S.H.,  
 RA LEFFLER H.;  
 RL J. BIOL. CHEM. 0:0-0(1997).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA GITT M.A., XIA Y.R., ATCHISON R.E., LUSIS A.J., BARONDES S.H.,  
 RA LEFFLER H.;  
 RL J. BIOL. CHEM. 273:2961-2970(1998).  
 DR ENBL; AF026799; G2914749;  
 DR ENBL; AF026796; G2914749; JOINED.  
 DR ENBL; AF026797; G2914749; JOINED.  
 DR ENBL; AF026798; G2914749; JOINED.  
 SQ SEQUENCE 301 AA; 34112 MW; 7C5AF97F CRC32;  
 Query Match 11.3%; Score 111; DB 11; Length 301;  
 Best Local Similarity 25.2%; Pred. No. 2.93e-03;  
 Matches 33; Conservative 36; Mismatches 56; Indels 6; Gaps 5;  
 Db 171 LPYVGLAGGFTVTRIIKGYVLPKTAFTAFNRVGSSEDIHLPRIQDCLVRNSYM 230  
 QY 171 LPYVGLAGGFTVTRIIKGYVLPKTAFTAFNRVGSSEDIHLPRIQDCLVRNSYM 230  
 Db 231 NGSGVTEERVAYN-PFGPGGFFDLSIRCGMDRPFVANGHILNFSRFPALRKINTLE 289  
 QY 231 NGSGVTEERVAYN-PFGPGGFFDLSIRCGMDRPFVANGHILNFSRFPALRKINTLE 289  
 Db 64 S-RWG-QKKLSAPFLFYQRFVEVLLIFQEGGLKALNGQGLGATSMNQALEQLREL 121  
 QY 64 S-RWG-QKKLSAPFLFYQRFVEVLLIFQEGGLKALNGQGLGATSMNQALEQLREL 121  
 Db 290 INGLDLSYVH 300  
 QY 290 INGLDLSYVH 300  
 Db 122 ISGVSQVLCVH 132  
 QY 122 ISGVSQVLCVH 132  
 RESULT 7  
 ID O44126 PRELIMINARY; PRT; 283 AA.  
 AC O44126  
 DT 01-JUN-1998 (TREMUREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMUREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMUREL. 07, LAST ANNOTATION UPDATE)  
 DE GALECTIN.  
 GN GAL-1.  
 OS HAEMONCHUS CONTORTUS.  
 OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; SPINURIDA.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-MOREDON;  
 RA NEWLANDS G.F.J., SKUCE P.J.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR ENBL; AF036098; G2687413;  
 DR PROSITE; PS00309; LECTIN\_GALACTOSIDE; 2.  
 SQ SEQUENCE 283 AA; 32530 MW; 9EA3746 CRC32;  
 Query Match 11.2%; Score 110; DB 5; Length 283;  
 Best Local Similarity 35.9%; Pred. No. 4.14e-03;  
 Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

[illegible]

Search completed: Tue Dec 15 15:56:10 1998  
Job time : 25 secs.